CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Mortgage Lending and Credit Risk: Micro-Level Data Analysis

Master's thesis

Author: Bc. Karolína Vachušková

Study program: Economics and Finance

Supervisor: doc. PhDr. Adam Geršl, Ph.D.

Year of defense: 2024

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants Charles University permission to reproduce and distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, April 29, 2024

Bc. Karolína Vachušková

Abstract

This thesis examines the effects of the debt service-to-income ratio (DSTI), debt-to-income ratio (DTI), and loan-to-value ratio (LTV) on credit risk. The dataset includes monthly loan-level data from around 250 thousand mortgages in the Czech Republic from July 2013 to July 2023. Using logit regressions, we confirm a positive effect of the level of DSTI, DTI and LTV on loan delinquency. Furthermore, we discover that the effects of these key lending variables are heterogeneous depending on the income and wealth classes as well as on the region of the client. Other explanatory variables align with general assumptions: a higher level of education, number of co-debtors, and GDP growth reduces the risk, whereas a higher interest rate increases the probability of delinquency. The thesis contributes to the debate on how effective macroprudential policy instruments using caps on DTI, DSTI and LTV are at employing a unique dataset.

Keywords

mortgage loans, credit risk, DSTI, LTV, microdata, Czech banking sector

Abstrakt

Tato práce zkoumá vliv poměru dluhové služby k příjmu (DSTI), poměru dluhu k příjmu (DTI) a poměru úvěru k hodnotě nemovitosti (LTV) na úvěrové riziko. Data zahrnují měsíční údaje na úrovni jednotlivých úvěrů z přibližně 250 tisíc hypoték v České republice od července 2013 do července 2023. Pomocí logistických regresí potvrzujeme pozitivní vliv výše DSTI, DTI a LTV na nesplácení úvěrů. Dále výsledky naznačují, že účinky těchto klíčových úvěrových proměnných jsou heterogenní v závislosti na příjmové a majetkové třídě, stejně jako na regionu, ve kterém klient žije. Ostatní vysvětlující proměnné jsou v souladu s obecnými předpoklady: vyšší úroveň vzdělání, počet spoludlužníků a růst HDP snižují riziko, zatímco vyšší úroková sazba zvyšuje pravděpodobnost delikvence. Diplomová práce přináší nové poznatky do diskuse o efektivitě nástrojů makroobezřetnostní politiky, které využívají stropy pro DTI, DSTI a LTV, a to na základě unikátního souboru dat.

Klíčová slova

hypoteční úvěry, kreditní riziko, DSTI, LTV, mikrodata, český bankovní sektor

Acknowledgments

The author is grateful to doc. PhDr. Adam Geršl, Ph.D. for his guidance, valuable comments and patience throughout the preparation and completion of this thesis.

Bibliographic Record

Vachušková, Karolína: *Mortgage Lending and Credit Risk: Micro-Level Data Analysis*. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 96. Advisor: doc. PhDr. Adam Geršl, Ph.D.

Contents

1 Introduction

During the last decades, a high emphasis has been placed on managing financial stability. One of the macroprudential policy instruments, which central banks and regulators use, is the so-called retail-lending parameters. Authorities set the caps on debt service-to-income ratio (DSTI), debt-to-income ratio (DTI), and loan-to-value ratio (LTV) to manage credit growth and subsequently control mortgage defaults. The DTI expresses the ratio of the amount of total indebtedness and the net annual income of the loan applicant. The DSTI is a percentage expression of the share of the average expenses arising from a consumer's total debt to net income. Lastly, LTV represents the ratio of a loan amount to the value of an asset purchased (CNB, 2024).

Several studies confirm the impact of these parameters on credit risk. The study by de Haan and Mastrogiacomo (2020) shows that the DSTI and LTV are positively associated with the probability of non-performance. Emekter et al. (2015) affirm a similar relationship between DTI and loan defaults. Campbell and Cocco (2015) prepare a theoretical model of mortgage default and conclude that the loosening of credit standards, such as higher LTV ratios, can increase the borrower's default risk. The size of DSTI and LTV caps and their effect on default is also investigated in the paper by Dietsch and Welter-Nicol (2014).

Our thesis aims to expand the debate by employing a unique dataset focused on the Czech banking sector. We formulate the hypotheses to evaluate (1) the effect of DSTI on household delinquency, (2) the effect of LTV on household delinquency, and (3) whether the significance and size of the effects differ across Czech regions.

The dataset contains loan-level data from one Czech commercial bank and covers approximately 250 thousand mortgages from July 2013 to July 2023. These data include originating DTI, DSTI and LTV, age of an applicant, and interest rate among other explanatory variables. Further, we also estimate the current DTI, DSTI and LTV. Besides borrower and loan-based characteristics, we complement the dataset with macroeconomic variables such as GDP (Gross Domestic Product) growth and unemployment rate. As the dependent variable, we create loan delinquency, where 1 represents loans 30 days past due or in default, and 0 otherwise.

Given the binary nature of this dependent variable, the thesis uses logit model to test the effects of retail-lending parameters. In the base model, we monitored the effect of DSTI and LTV at origination and compared these relationships across different subsets. These subsets include the division of loans into groups according to the size of DSTI and LTV, Czech regions, amount of initial income, and property type. The results confirm our hypotheses: both DSTI and LTV significantly affect household delinquency, and these effects differ across regions. Furthermore, we discover that the effects of these key lending variables are heterogeneous depending on the income and wealth classes. However, there is still room for further research in the area of current parameters, as the current parameters are estimated.

Since the dependent variable is imbalanced, we employ two methods that penalise the model for overfitting the majority class. According to Woo et al. (2022), we use the Firth and the RE logit models, which are suited for a rare event data. These validations ensure that the results of the previous models are credible without bias. For further robustness check, we construct two more dependent variables: the first binary variable monitors non-performance, and the second ordered variable corresponds to one of three CNB (Czech National Bank) rating categories. The effects of originating DSTI and LTV remain largely unchanged, and we affirm that our evaluation of the hypotheses still holds.

The thesis is structured as follows. Chapter 2 reviews the relevant literature on credit risk and retail-lending parameters. Chapter 3 focuses on the description of data used in the analysis. Chapter 4 presents the methods employed to estimate our models and states the hypotheses. Chapter 5 provides regression results. Finally, Chapter 6 summarises the main findings and conclusions.

2 Literature Review

This chapter contains several subsections. The first subsection introduces papers relevant to this thesis that use similar methodology and focus on credit risk and retail-lending parameters. The second subsection focuses on retail-lending parameters as the macroprudential policy instruments. The last part presents the methodology specification.

2.1 Credit Risk and Retail-Lending Parameters

The following studies concentrate on credit defaults, among other factors at the microdata level, highlighting the effects of the limits on mortgage loans. Most of the papers, which works with logistic or probabilistic regression models, offer insights into potential expected outcomes, and provide valuable guidance on incorporating explanatory variables.

The paper by de Haan and Mastrogiacomo (2020) uses a logit probability model on loan-level data to analyse the factors that contributes to the non-performance of mortgages in the Netherlands between 1996 and 2015. According to the author's expectations, the study finds that the originating loan-to-value (OLTV) ratio and the debt service-to-income (ODSTI) ratio are positively associated with the probability of non-performance. The analysis uses marginal effects and shows that the non-performing probability increases by 0.19 percentage points with ten percentage points higher OLTV, respectably by 0.75 percentage points with ten percentage points higher DSTI. Besides OLTV and ODSTI, other explanatory variables were loan type, borrower age, the applicant's self-employment status, and the Netherlands' specification: the so-called Nationale Hypotheek Garantie, which are loans with government-backed debt insurance. This guarantee enables us to split off the mortgages into two groups - insured and uninsured. The study concludes that mortgages with government loan guarantees perform better and that certain loan and borrower characteristics, such as the loan type or the underwater position of the collateral, increase credit risk. The authors suggest that the OLTV limit should be set to about 70-80% for uninsured mortgages and 90% for those with mortgage insurance to avoid acceleration of non-performance probabilities. The findings of this study can be useful for policymakers and lenders in designing effective mortgage policies and managing credit risk.

Saha et al. (2022) introduce a study using micro-level data between 2001 and 2017 from an Indian public sector bank. The analysis is specific by addressing the issue of endogeneity in the LTV ratio while analysing the drivers of housing loan default. To deal with the endogeneity in their probit regression, the authors introduce two instrumental variables: the bank's cash reserve ratio of the loan's originating year and the risk weight for the lowest strata of mortgage loans. A comparison of models' estimations confirms that results would be biased without the IVs. Again, as the authors expected, higher LTV positively affects the default probability. Apart from the LTV ratio, the other key determinants of default are the interest rate, frequency of repayment, prepayment options, and loan period. The reduction in probability stems from an increase in per capita income and an increase in the number of employed people in the state, which both reflect borrowers' ability to pay by borrowers. The study also includes control variables, including the change in the IRB reclassified default in 2014 from 180 days to 90. This change proves to be not statistically significant. The study's findings provide valuable insights for banks and housing finance companies to formulate appropriate strategies for lending to this sector.

The paper by Gaudêncio et al. (2019) focuses on lending standards and their influence on default rates in the euro area. To analyse the characteristics of residential real estate loans from 2003-2018 the authors use loan-level data from the European Datawarehouse. This type of dataset allows authors to do cross-country comparisons. In contrast, they need to consider country-specific heterogeneity in their probit model. Results confirm that the key impact of lending standards are LTV and loan-to-income (LTI) ratios at origination, original loan maturity, and borrower employment status. Particular outcomes are calculated using marginal effects (at means): a 10 percentage point increase in OLTV increases the default probability by 0.2 percentage points, and a 10 percentage point increase in OLTI raises the default probability by 1 percentage point.

One of the few papers that study DTI is the paper by Emekter et al. (2015). The authors utilise logit model to evaluate credit risk in online Peer-to-Peer lending on loan-level data. They conclude that borrowers with higher FICO scores (a score that assess an applicant's credit risk), better credit grades, lower DTI ratios, and lower revolving line utilisation are associated with a reduced risk of default.

Agbemava et al. (2016) study Ghana's microfinance institutions. Their logistic regression model predicts the probability of loan defaults based on the selected independent variables. The results show that six factors were statistically significant; these factors are marital status, dependents, type of collateral or security, assessment, duration, and loan type. The predicted default rate based on the model is 86.7%. The researchers provide several recommendations on how to improve institutions' risk assessment and methodology.

Another logit model is presented in the study by Kelly and O'Toole (2018), who estimate the probability of default as a function of OLTV and original rent coverage (ORC). ORC is the ratio of the monthly rent to the monthly mortgage payment and serves as a proxy for the debt-service ratio. On a loan-level dataset from the UK buy-to-let market, the researchers find that loans with an OLTV of more than 75 and an ORC under 1.5 were associated a large increase in default risk.

The paper by Gerlach-Kristen and Lyons (2018) chooses the random effects ordinary least squares (OLS) model to study the mortgage arrears in European households, although the dependent variable is a dummy variable. The examination is completed with comprehensive robustness tests. The panel dataset contains 15 European countries from 2004 to 2011. The authors focus on affordability problems, which spring from unemployment or high mortgage payments, and negative equity (resulting from a house price fall). Results suggest that while affordability problems matter for both temporary and long-term arrears, the negative equity is insignificant for both groups. The study explains the negative equity insignificance by European recourse legislation. A combination of these variables, double trigger, turns out to be positively significant for those with long-term arrears issues. Finally, the arrears are more common in countries with lower incomes and fewer investor protections.

Elul et al. (2010) analyse the drivers of mortgage default, specifically negative equity and illiquidity. As a dependent variable in their dynamic logistic model the authors use a dummy variable, whether the loan is 60 or more days after payment. The examined US loan-level mortgage data includes credit bureau information, which allows us to take into consideration the combined loan-to-value ratios. Calculating combined LTV shows that for over one-quarter of borrowers with two mortgages, using only the LTV of the first mortgage would underestimate the risk by 15%. The second mortgage also significantly increases the default risk, especially for the debtor with the first mortgage LTV close to 100%. Negative equity and illiquidity appear to have similar significant effects on mortgage default, and above that, these drivers interact with each other. The influence of illiquidity on default rises even with low combined LTV. Finally, the country-level unemployment shocks are also linked with increased default risk and interact substantially with combined LTV.

Quercia et al. (2012) use a logit model within the comparison of low- and moderate-income households in terms of the likelihood of mortgage default and prepayment among households. The study concludes in line with expectations: lower-income households had a higher probability of default and a lower probability of prepayment, even within the moderate and low-income categories.

Further paper that uses logistic regression and works with micro-level is Rahman (2013), which determines characteristics contributing to the probability of a household being poor. The research data is collected in Bangladesh between 2008 and 2009. Significant factors that contribute to poverty are occupational status, education, household size, dependency, and the fact that women earn less than men for the same work.

On microdata from Statistics Finland, Herrala and Kauko (2007) develop a microsimulation model that explores the relationship between household distress and macroeconomic shock. This topic is usually studied using aggregated data; the inclusion of microdata aims to match loan-based characteristics at the household level and look at financial fragility more precisely. The paper considers three potential shocks: unemployment, housing prices, and interest rates. The dependent variable in the logit model is whether the household is financially distressed or not . The independent variables include net income after tax and loan servicing costs, as well as a vector of potentially useful control variables included in the data set. The conclusion is that only under the most extreme scenario might the household credit risk jeopardize financial stability.

Another work employing some kind of shock is Froyland and Larsen (2002). They examine the economic response to house price shock by conducting two stress tests. Thanks to the solid OLTV of Norwegian household loans, even after a fundamental fall in property prices, a substantial part of collateral value is maintained above the value of the loan. The fall, however, reduces the household consumption and housing wealth , leading to higher losses for financial institutions.

A commonly covered topic is the link between LGD and LTV. The drivers of Loss Given Default (LGD) are studied by Qi and Yang (2008). Their research contains data from the entire housing market cycle and examines the US high-LTV mortgages that have defaulted from 1990-2003. Between the explanatory variables besides LTV are details about the default and property characteristics. According to the regressions' results, the current LTV is the single most significant determinant of LGD and a much more accurate predictor for LGD than the original LTV. Greve and Hahnenstein (2014) demonstrate that the structure of the underlying LTV distribution has a significant impact on how stress affects the average LGD of a Real Estate loan portfolio.

Only a few studys have worked with such a detailed loan-level dataset in our demographic area. One frequent source for the Czech banking sector is internal data from the database of the CNB's Central Credit Register (CRC). This database has harvested all credit relationships between companies and banks in the Czech Republic since 2002 with a monthly frequency. Geršl et al. (2015) use a loan-level dataset for non-financial firms and focus on the impact of monetary conditions on the risk-taking behaviour of banks in the Czech Republic. Geršl and Jakubík (2011) published the first analysis using the CRC's data focusing on firms' financing and their lending relationship in the Czech Republic. Another dataset based on the internal CNB database is presented in the paper by Derviz and Raková (2012). They collected data about the ten biggest Czech commercial banks based on a loan-level dataset of newly granted loans to non-financial businesses.

2.2 Macroprudential Policy and Retail-Lending Parameters

Macroprudential policy instruments have recently become a popular area through which central banks and regulators are trying to manage credit growth. Part of these instruments includes setting caps on retail-lending parameters. ESRB (2021) provides specific examples, e.g., in 2019, about half of European banks used DSTI limits. LTV limits were even more common, by over two-thirds of the countries. In response to COVID-19, some banks, including the CNB, lowered these limits to stimulate the economy. Overall, these instruments rank among the most extensively employed.

Literature assesses the macroprudential instruments mainly as efficient. Gross and Población (2017) introduce a macro-micro model with household-level data focusing on four European countries. In its time, the model was unique for assessing the impact of LTV and DSTI caps on households' risk behaviour. The results were in line with the expectations; through all observed countries, the model confirms that LTV caps are related and help decrease LGDs; the same could apply to the relationship between DSTI caps and PDs. Moreover, the researchers find cross-risk effects when PDs react on LTV caps and LGDs on DSTI caps. This relationship appears to be stronger for DSTI caps.

This work is extended by Jurča et al. (2020) by introducing an endogenous loan-granting feature. Thanks to this feature, the model considers the option of banks adjusting their lending standards in response to changes in macroeconomic conditions and borrower characteristics. Based on data from Slovakia, it was proved that a combination of borrowed-based measures (DTI, DSTI, and LTV) clearly improved the resilience of both banks and clients.

Another insight into the relationship between credit risk and LTV, DSTI or a combination of both measures brings Dietsch and Welter-Nicol (2014). The paper works with approximately 850 thousand French individual housing loans, including loans and borrowers' characteristics; this dataset was conducted during the first decade of 2000, and the extent makes it unique. According to the study, the highest credit risk is concentrated not in the tranches with the highest LTV and DSTI but at an LTV level of around 95%, respectively at DSTI level around 33%, which are standard used caps. The explanation of this non-linear relationship is that the borrowers in highest tranches are usually in upper income and wealth classes. Considering this French example, the regulatory requirements are above the need of the worst-case scenarios caused by excessive LTV or DSTI ratios.

Campbell and Cocco (2015), with their theoretical model of mortgage default, provide complete insights into the factors that influence mortgage default decisions and the implications for mortgage lenders and policymakers. The paper distinguishes between adjustable and fixed mortgage rates, while borrowers with ARMs are more likely to default during a housing downturn. Besides other conclusions, the authors note that the loosening of credit standards, such as higher LTV ratios, can increase the risk of default for borrowers.

2.3 Specification of Methodology

Working with credit defaults can be challenging if the examined portfolio is low default. Penikas (2020) defines the low default portfolio as a portfolio where the number of defaults is less than 25, or the default rate is less than 3% of total observations. According to Pluto and Tasche (2011), this issue is typical for highly reliable clients' portfolios (e.g., sovereign) or portfolios with high-volume exposure (e.g. specialized lending). Several different methods for estimating probabilities of default for this type of portfolio are tested by Dzidzeviciute (2012).

Several papers focus on binary choice models for rare events data. The paper by Jin et al. (2005) prepares a score test which guide in choosing between logit and probit models in such cases. Both models showed similar coefficient signs and magnitudes, but significant differences in predicted probabilities, where logit model performs better. Woo et al. (2022) compare not only probit and logit models, but also Firth logit and RE logit models. The Firth model applies a penalization to the likelihood which helps to eliminate bias from the maximum likelihood estimator (Firth, 1993). Further, the Rare Event (RE) model uses weights to each observation in the dataset to balance the influence of rare events (King and Zeng, 2001). Both methods offer substantial improvements over probit and logit models, and the paper conclude with a guide on how to choose the correct approach based on the data characteristics.

3 Data

This chapter presents the data in our analysis both verbally and graphically. The first section describes the dataset and dependent and explanatory variables. The second part offers descriptive statistics.

3.1 Data Description

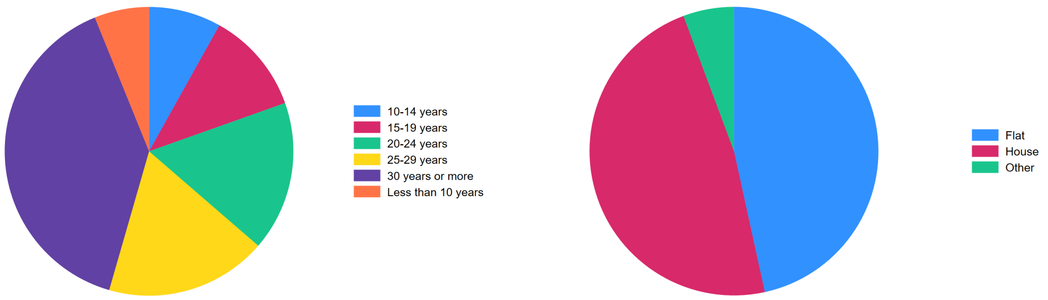
The loan-level data are harvested from one Czech commercial bank.[1] The dataset contains monthly data from around 250 thousand mortgages from July 2013 to July 2023. A range of macroeconomic variables (i.e., GDP growth and unemployment rate) complements the dataset. These macroeconomic data are downloaded from CNB and Czech Statistical Office (CZSO).

Introducing the properties of our dataset, the first chart 1a illustrates the variety of mortgage terms that applicants choose at origination, with a significant portion taking a long-term commitment of 30 years or more. Further, the second figure 1b reveals a clear preference for certain types of property when houses and flats covered most of the mortgage collateral. The *Other* category includes lands, recreational properties, garages, and more. All examined mortgages are in CZK, which eliminates the exchange rate risk. Both the primary mortgage applicant and the property type are from the Czech Republic.

In the loan-level dataset, missing values occurred for various reasons. We deleted all the bugs that contained such values before the final modelling. Because of earlier variations in data collection and structure, the dataset contains mortgages originating in 2011 and later. Therefore, we can deduce the dataset does not include mortgages that would have been affected by the Global Financial

[1]Please note that this thesis works with internal, non-public data; as such, a full disclosure of all details and specific levels of examined variables is not possible.

Figure 1: Mortgage Distribution



(a) by Loan Maturity (b) by Property Type

Note: The figures are based on data gathered during a mortgage application.

Crisis. As mentioned above, we focus on the Czech Republic; thus, we excluded all mortgages with an applicant or collateral property outside the Czech Republic. Further adjustments and data cleaning details are mentioned in the descriptions of individual variables.

3.1.1 Dependent Variables

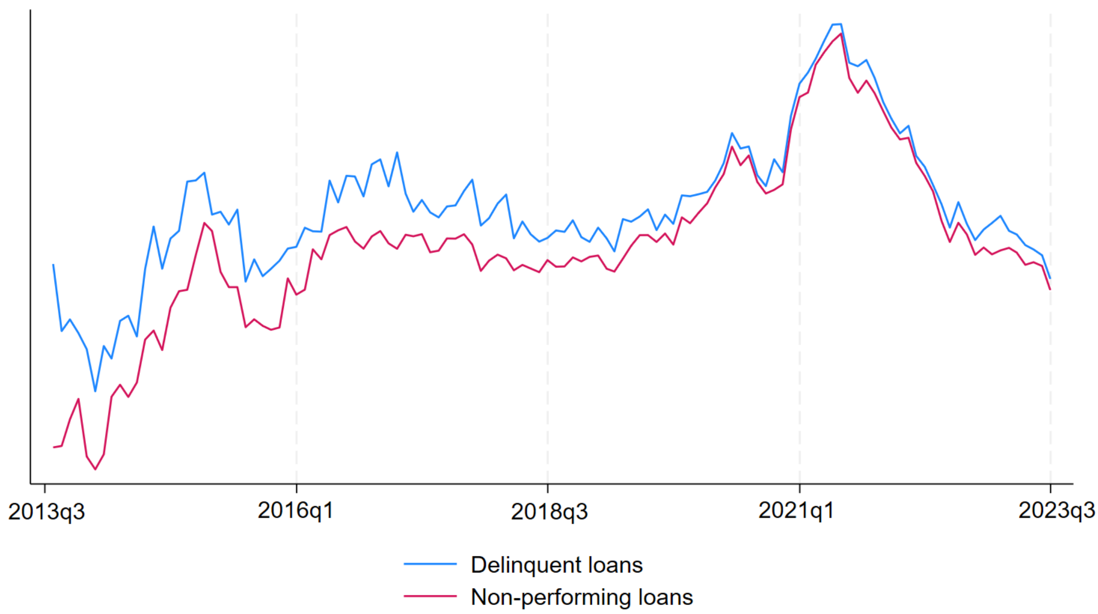
We are interested in loan repayment as a dependent variable for our analysis. Inspired by Elul et al. (2010), who utilized a dummy variable to indicate loans 60 or more days overdue, we used a 30-day past-due benchmark.

Most papers in this area focus on loan performance, for instance, de Haan and Mastrogiacomo (2020) or Saha et al. (2022). The BCBS (2016), aligning with the IFRS 9 standards, states the following identification of non-performing loans:

1. loans classified as defaulted following the Basel framework,
2. loans classified as impaired under the applicable accounting framework, and
3. loans that are not defaulted or impaired but either
   (a) are 90 or more days past due, or
   (b) there is evidence it is unlikely that repayment will be fulfilled without the realisation of collateral.

However, during the observed years, the Bank's mortgage portfolio displayed a very low frequency of non-performing loans. Therefore, we've expanded our focus on non-performing loans to cover loans 30 days past due. The final variable is called *Delinquent Loan*, where 1 indicates mortgage payment is 30 or more days past due or non-performing and 0 otherwise. Despite this extension, our binary dependent variables show rare event rates. Non-performing loans remain within our scope: a dependent variable focused on loan-performance is used as a robustness check (see Chapter 6). Figure 2 compares the development of delinquent and non-performing loans.

Figure 2: Development of Delinquent and Non-performing Loans

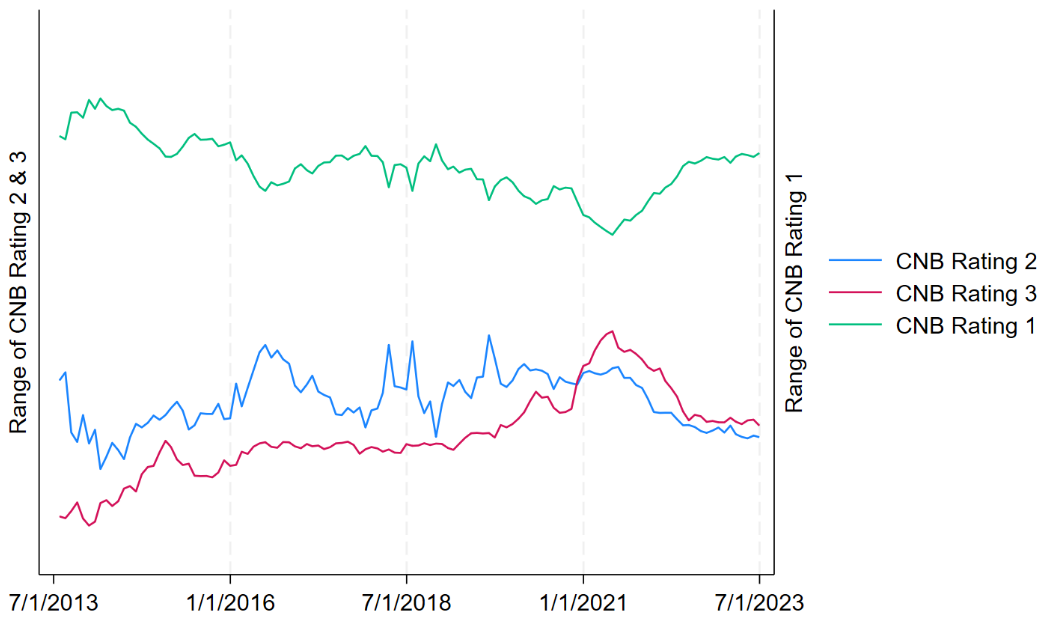


Please note: Specific values cannot be disclosed; the figure shows trends only.

The *Delinquent Loan*, serving as the dependent variable in our models, is applied with a one-month shift ($y_{t+1} \sim x_t$). By implementing time shifts, we align the independent variables with the temporal changes that lead to clients' non-payment. To avoid the occurrence of null observations, we shortened the dataset's observation period by a month, examining from July 2013 to June 2023. Different time shifts and various sub-datasets are detailed in Chapter 6.

The final variable used for the robustness check is CNB's loan classification. The CNB's rating ranges from 1 to 5, with category 1 meaning standard, 2 watch, 3 nonstandard, 4 doubtful and 5 loss loans (CNB, 2003). We created a dependent ordered variable, where 1 corresponds to CNB 1 (*standard* category), 2 corresponds to CNB 2 (*watch* category), and 3 corresponds to CNB 3-5 (*non-performing* category). Figure 3 displays the development of these three variables over time. The categorical variable *CNB Rating* is unique for being ordered, as opposed to the binary nature of the other dependent variables.

Figure 3: Development of CNB Rating



Note: The ratio of $CNB = 2$ and $CNB = 3$ is depicted on the left axis, while $CNB = 1$ is represented on the right axis.

Banks applied the CNB's loan classification before IFRS 9 loan staging: The IFRS 9 established three stages of impairment assets. Stage 1 is the category for standard-performing loans with no or very low change in credit risk from their initial recognition. Banks calculate the provision based on 12-month expected credit loss (ECL). As the credit risk increases, the loan moves to stage 2, where the basis for calculating the provision is the lifetime expected loss over. Within stage 2, the loan is still performing. Finally, non-performing loans are in stage

3. Recognition of ECL is the same as in stage 2. However, while interest income for a performing loan is calculated on the gross carrying amount, stage 3 interest income uses a net carrying amount (BIS, 2017).

The loan staging occurred in 2018, and this information cannot be collected for the whole dataset, rather only for the last several years. The other reason for being careful with staging occurred during the COVID-19 crisis. After the announcement of the moratorium on the repayment of loans and mortgages, some banks temporarily transferred these clients to stage 2 for preventive reasons (Fio, 2021). This transfer could affect the results of the models. Fortunately for our analysis, the Bank collects the information about CNB's classification even after the introduction of IFRS 9 staging. We can roughly consider the three CNB's categories, we presented, as an estimate of today's IFRS9 stages.

We also considered an approach in which the thesis would use the development of the loan's Probability of Default (PD)[2] as another dependent variable if the occurrence of defaults proved to be statistically insignificant. This approach means that the dependent variable would represent a probability ranging from 0% to 100%, with 100% indicating a default. To maintain the binary nature of a dependent variable, we intended that loans maintaining the same or a lower one-year PD compared to the previous period be coded as 0, while loans with a worsening one-year PD or those in default be coded as 1. However, this concept was discarded for two primary reasons. The first is that the methodology for calculating PD before and after the implementation of IFRS9 is inconsistent, thus making any comparative analysis inaccurate. The second reason is that some variables used in modelling PD are also included as explanatory variables on the right side of our models. Using PD as a dependent variable would introduce bias into the results and distort the validity of the econometric analysis. A similar reason for

[2]The Probability of Default (PD) is the probability that a borrower will not meet their repayment obligations over a specific period.

not including it in the analysis also applies to Loss Given Default (LGD).[3]

3.1.2 Explanatory Variables

Key explanatory variables for our analysis include DTI, DSTI, and LTV. The caps for these indicators might be established to manage risks associated with providing retail-secured residential property loans. The European Systemic Risk Board (ESRB) has classified these limits as macroprudential tools, which supervisory authorities can employ to enhance financial stability beyond the CRD IV directive's tools (ESRB, 2013). While DTI and DSTI caps regulate the size of debt relative to household income, the LTV cap regulates the maximum amount that can be lent to the borrowers against their underlying collateral. These instruments are commonly used across European countries (ESRB, 2021). In the Czech Republic, the central bank initially implemented LTV limits in 2015 and subsequently introduced DTI and DSTI levels in 2018. From January 1, 2024,[4] the CNB only requires the LTV cap of 80%, or for applicants under the age of 36 years 90% (CNB, 2024).

The bank collects these metrics when applying for a loan; unfortunately, not all data is regularly updated. Given this limitation, we made specific choices to maintain data relevance and accuracy as much as possible.

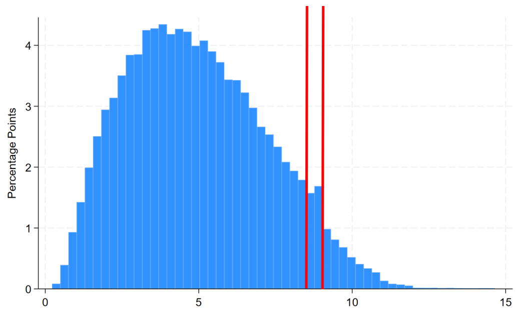
Firstly, we retained the originating DTI (ODTI), originating DSTI (ODSTI) and originating LTV (OLTV), which represents the applicant's metrics at the time of their mortgage application. Due to their accuracy, we consider these variables as fundamental. Nevertheless, we cleaned this data of outliers or erroneous values (i.e., values smaller than 0). See the distribution of ODTI, ODSTI and OLTV in Figure 4 below. Beyond the distributions, we complement the figures with

[3] Loss Given Default (LGD) is an estimate of credit loss a bank might suffer when a borrower defaults on a loan.

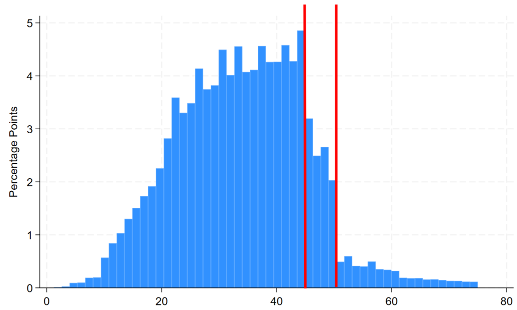
[4] The current credit ratio upper limits set by the CNB are accessible here: Requirements for LTV, DSTI and DTI limits - Czech National Bank (cnb.cz).

CNB parameters caps: DTI caps were set between 8.5-9, DSTI caps ranged from 45-50%, and LTV caps varied from 80-100%. See footnote 4 for details.

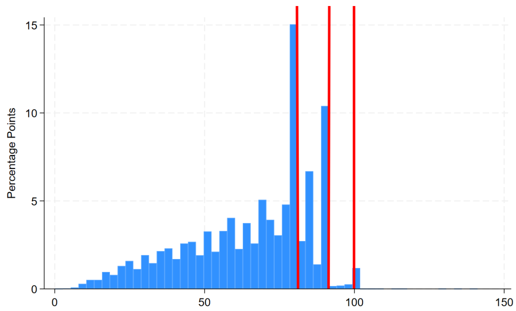
Figure 4: Distribution of Parameters at Mortgage Origination



(a) ODTI



(b) ODSTI



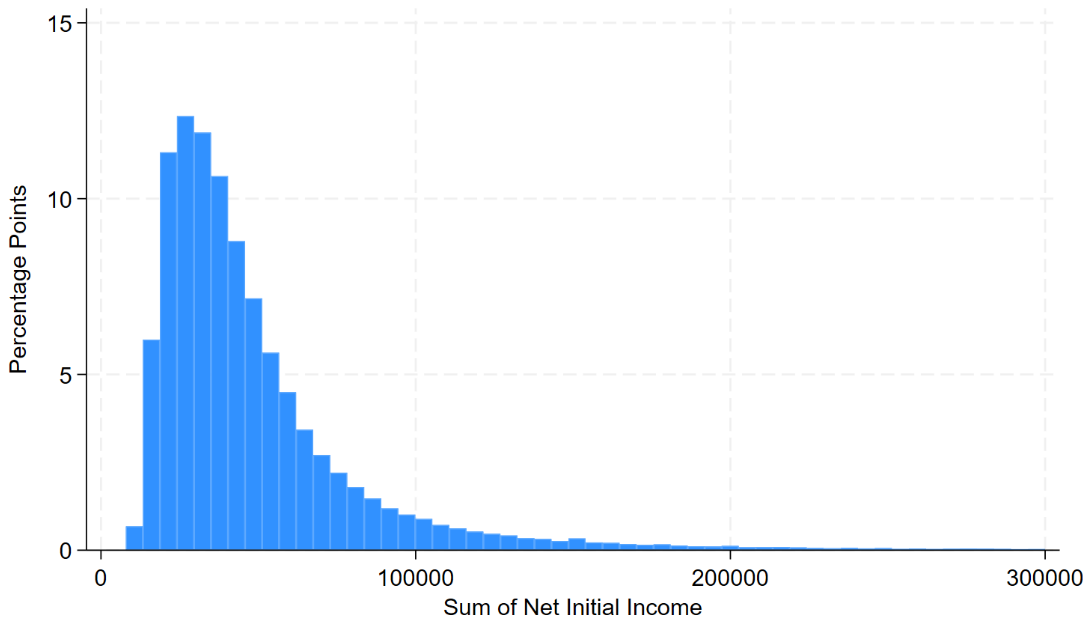
(c) OLTV

Note: The jagged appearance of the distribution is probably caused by the set caps required by the CNB and the combination of different sources when collecting this data (various sources in the bank have different number of decimal places and variables are sometimes given as percentages and sometimes as ratios).

Secondly, to capture the current financial situation of borrowers, we computed the current DTI, DSTI and LTV values. The current DSTI (CDSTI) was recalculated using the most recent income data available from the bank; in cases where updated income data was unavailable, we adjusted the last known income by the rate of nominal wage growth. For this operation, we used annual data from the CZSO on the average gross salary. We conducted the income updates once a year, based on the assumption that the majority of applicants' salaries do not increase continuously each month but rather on an annual scale. The numerator reflects the current total repayment obligations of the borrowers, including not only the repayments for the mortgage but also other known existing loan obligations. From the calculation (and from the dataset) we excluded mortgages for which a one-time repayment was assumed, and applicants with an initial salary exceeding 1 million CZK (less than 0.01% of applicants).

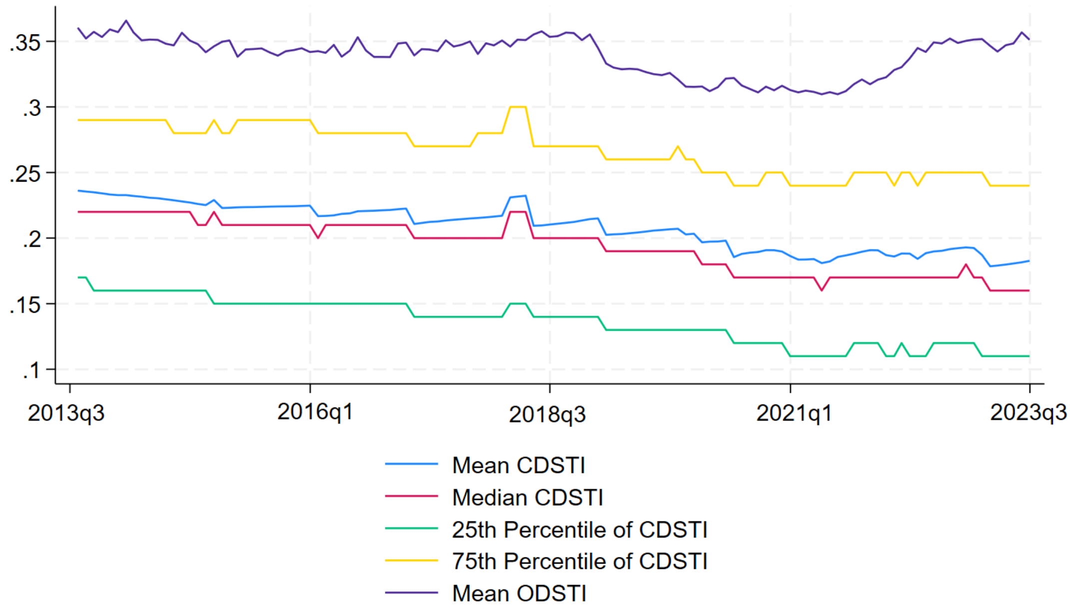
Figure 5 presents the distribution of total initial income for a loan with a detailed distribution of income below 300 thousand CZK (income over 300 thousand CZK has less than 1% of loans). Figure 6 shows the average and quartile distribution of the calculated CDSTI over time. The mean and median ratios, along with the 25th and 75th percentiles, indicate a consistent range of debt service burden among borrowers. Figure 6 is complemented by the average DSTI at origination in any given month.

Figure 5: The Detailed Distribution of Net Initial Income Below 300 Thousand CZK



To obtain current DTI (CDTI), we utilize the current income, which we adjusted before. The monthly income was multiplied by 12 to derive the annual income. The numerator corresponded to the remaining debt value of the mort-

Figure 6: Development of DSTI



Note: The jagged pattern in the Figure is a result of the annual salary update.

gage.[5] Find development of DTI in Figure 7.

Similarly, we recalculated the current LTV (CLTV) ratio. In most cases, the bank has annually updated property value data; these were directly used to determine the denominator of the CLTV ratio. When the property value was unchanged, we modified the last known property price by the Housing Price Index (HPI), utilizing the annual growth rate obtained from the CNB. The numerator was derived from the current debt value of the mortgage. The consistent pattern of LTV ratios in Figure 8 indicates a stable level of risk and equity within the examined mortgage portfolio.

Additional loan-level data, which are collected at origination and in principal stable, include demographic details that help to understand the profiles of mortgage applicants. The data is visualized in Figure 9: the first chart captures the

[5] In contrast to CDSTI, where we consider total repayment obligations

Figure 7: Development of DTI

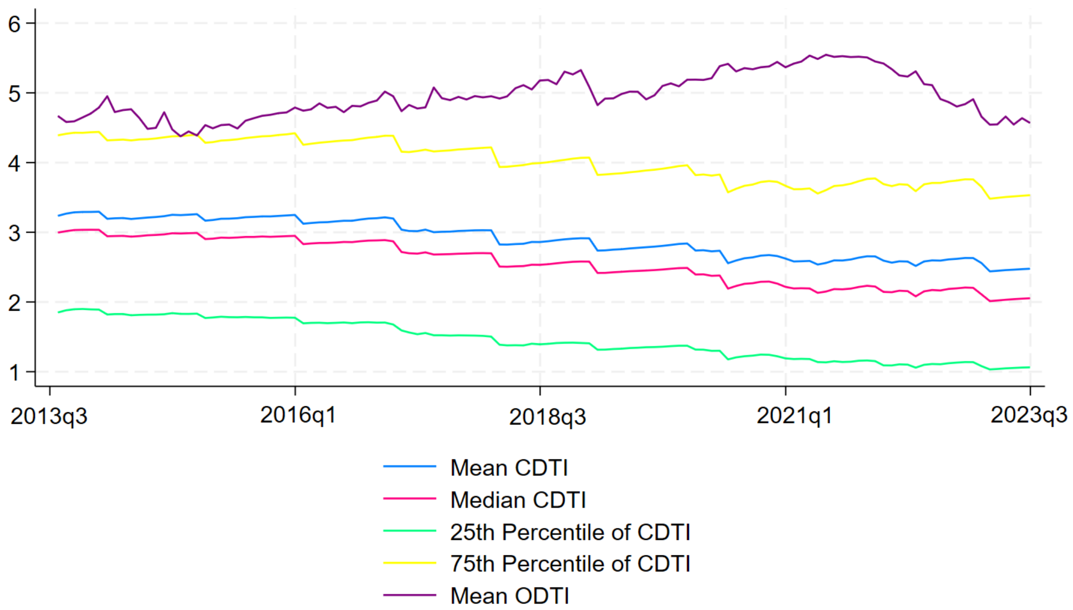
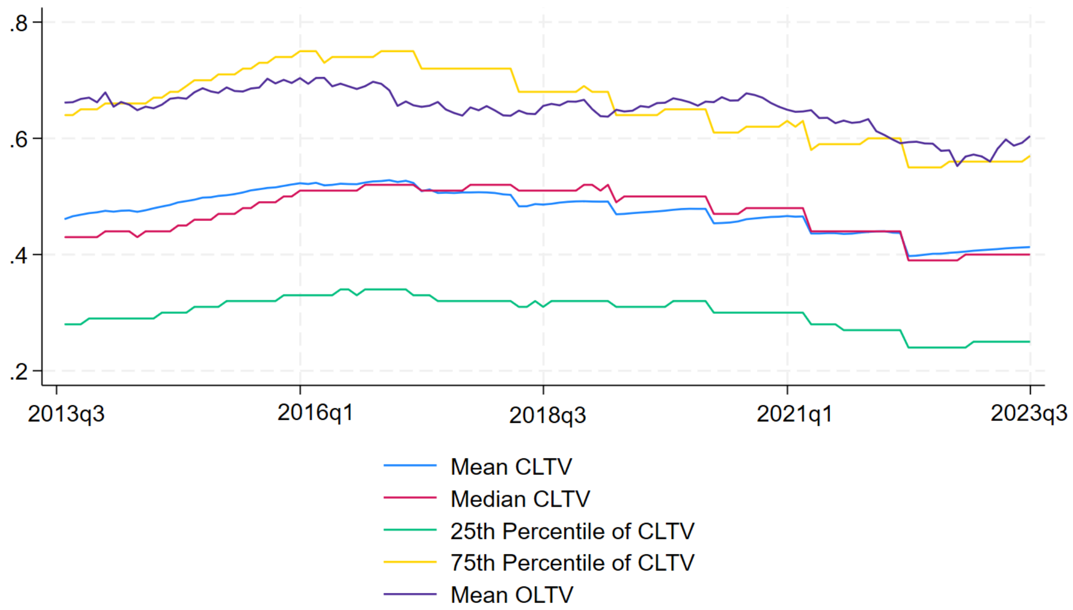
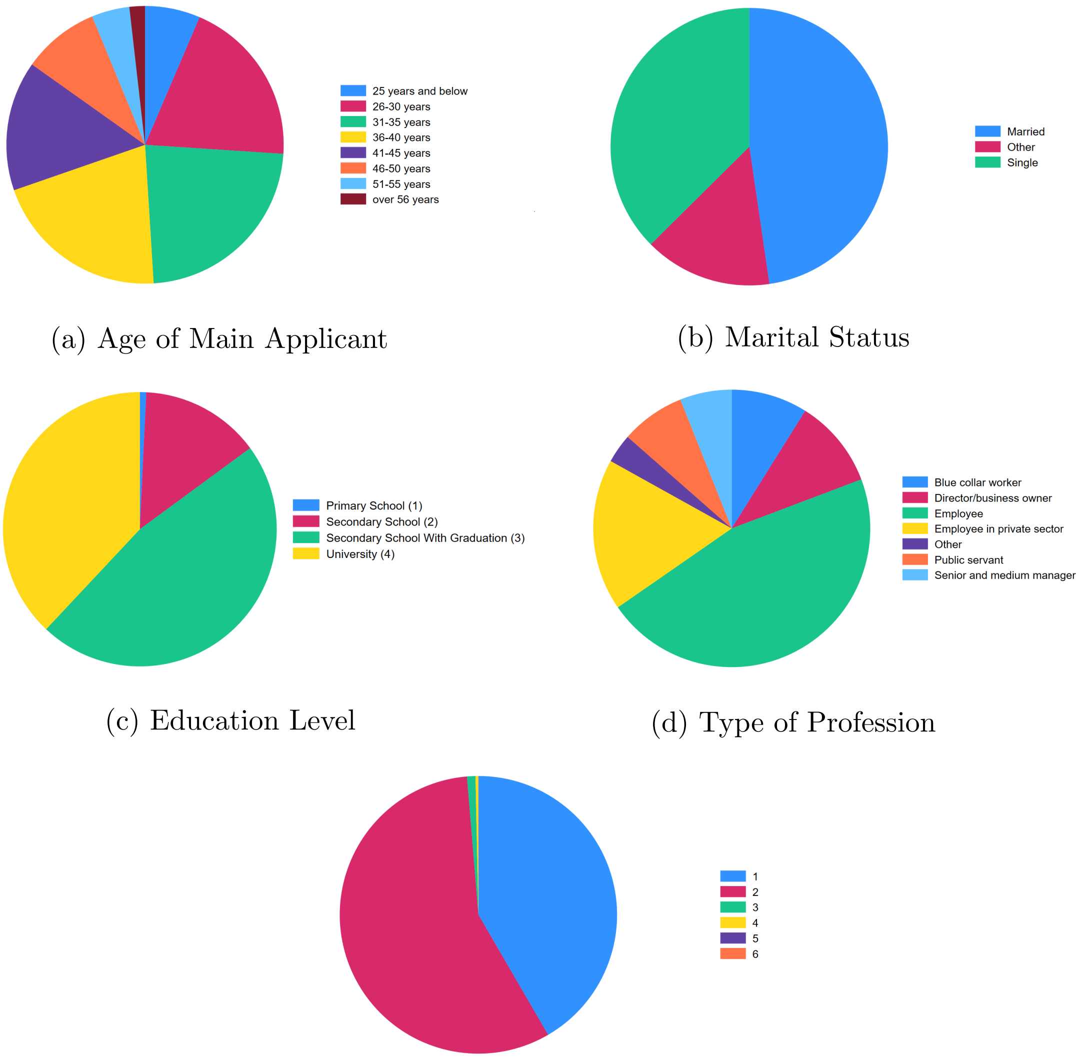


Figure 8: Development of LTV



age distribution, from those 25 years and below to those over 56 years, mapping out the age-related tendencies of those seeking mortgages. The second chart portrays the marital status diversity, where besides *Married* and *Single*, there is a category *Other* for divorced, widowed and others. The third figure presents the educational

Figure 9: Demographic and Other Details



(a) Age of Main Applicant

(b) Marital Status

(c) Education Level

(d) Type of Profession

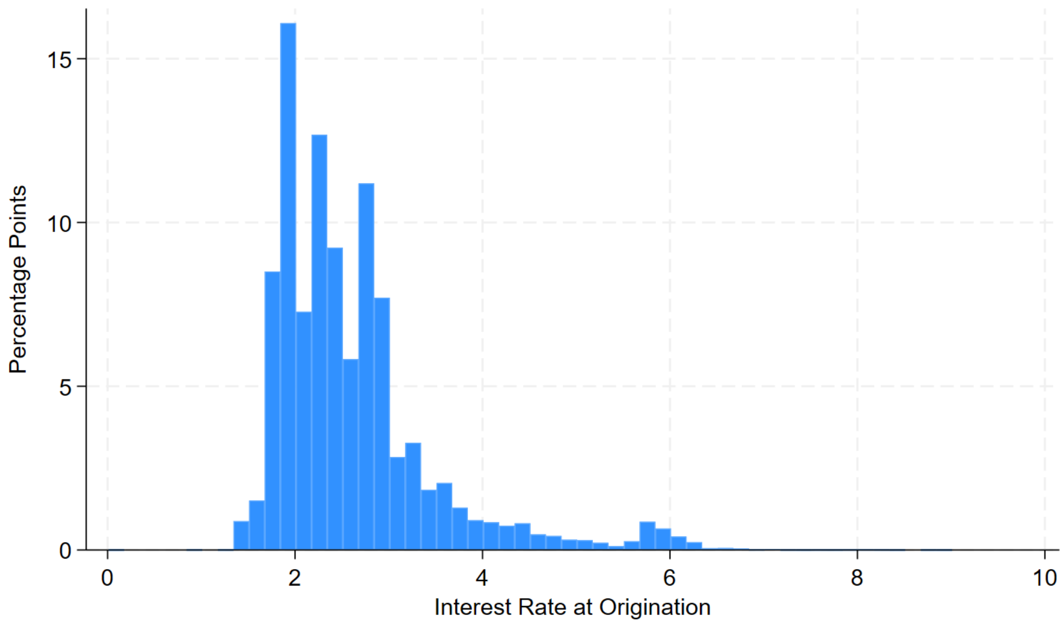
(e) Number of Debtors

Note: Data at mortgage origination.

backgrounds of applicants ranging from primary school through to university graduates. The variable is numeric, where a value of 1 corresponds to applicants with a primary school education, and subsequent values representing higher educational levels. We removed loans from the dataset for which the main applicant's educational background was not specified. The fourth chart categorizes applicants by profession; nearly two-thirds of all borrowers are employed. The final figure shows the number of debtors at the time of application. Further, for each mortgage, we collected details on both the region of the applicant's permanent residence and the region where the loan's property is found.

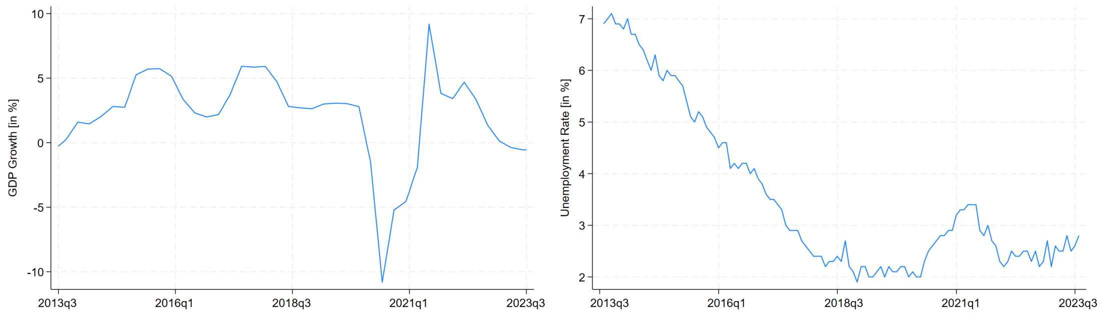
Loan-based characteristics, which evolve over time, encompass interest rate, mortgage principal, open, close and the actual mortgage maturity date. To the final dataset, we further incorporated a variable representing the mortgage age, calculated as the difference between the current month and the month in which the mortgage was open. However, it is important to note that this variable is limited by the fact that the majority of mortgages in the dataset do not undergo the entire standard repayment process since 86% of the mortgages have a maturity exceeding 15 years, and the oldest mortgages were initiated in 2011. Figure 10 depicts the interest rate at loan granting; on average, the rates began to rise in the second half of the observed period, which reflects the development of interest rates set by the CNB.

Figure 10: The Distribution of Interest Rate [in %] at Loan Granting



The macroeconomic variables in our analysis consist of GDP, unemployment, and indirectly also the growth of wages and HPI. The GDP data is sourced from the CNB database and comprises quarterly, seasonally adjusted figures. For months where data was missing, we calculated the value based on the previous and following quarter to smooth the data. The CZSO provided monthly seasonally adjusted unemployment data. Further, as mentioned earlier, the calculation of DTI, DSTI and LTV incorporated the annual growth in nominal wages from the CZSO and the HPI, which is an annual measure available in the CNB database.

Figure 11: Development of GDP Growth & Unemployment Rate



(a) GDP Growth (b) Unemployment Rate

In an effort to improve the analysis' specificity, we attempt to find regional data for these variables. This data would have allowed us to match macroeconomic conditions more precisely to individual loan applicants based on their addresses. Unfortunately, we found that the CZSO collects regional data only on an annual basis, and the most recent available data was for the year 2021, which was not sufficient.

3.2 Descriptive Statistics

This section outlines the descriptive statistics for the key variables in our models, as shown in Table 1.

At the top of the table, we observe the results of our dependent variables. Comparing the delinquent and non-performing loans, it is apparent that the former has a marginally higher mean, indicating a greater presence of clients with a 1 rating in the dataset. Regarding the CNB rating, which varies from 1 to 3, the data confirms that the standard rating is the most common, with the mean calculated at 1.0082.

Table 1: Summary Statistics

Variables	Mean	Median	Min	Max	Standard deviation
Delinquent Loans	0.0029	0	0	1	0.0537
Non-performing Loans	0.0028	0	0	1	0.0526
CNB rating	1.0082	1	1	3	0.1149
ODTI [in %]	4.8	4.6	0.2	14.7	2.2943
CDTI [in %]	2.8	2.5	0.0	14.7	1.9112
ODSTI [in %]	34.4	34.2	1.2	75.0	11.6758
CDSTI [in %]	20.2	19.0	0.0	75.0	9.7627
OLTV [in %]	66.2	72.0	0.1	141.1	21.3474
CLTV [in %]	46.6	46.0	0.0	141.1	23.1570
Education Level of Main Borrower	3.2	3	1	4	0.7197
Current Interest Rate [in %]	2.4	2.3	0.01	10.1	0.6078
Number of Debtors	1.6	2	1	6	0.5301
Current Age of Main Borrower	39.6	39	18	78	8.3265
Current Age of Loan [in years]	3.3	2.8	0.0	12.3	2.4318
GDP [in %]	2.0	2.8	-10.8	9.2	3.5485
Unemployment [in %]	3.9	3.4	1.9	11.5	1.5544

Focusing on retail lending conditions at the time of loan origination, ODTI shows an average of 4.8% and a median of 4.6%, with values ranging from 0.2% to a high of 14.7%, suggesting diverse debt-to-income circumstances at the loan's initiation. The lower CDTI mean and median values reflect a decrease in debt relative to (mostly increasing) income over time. A similar assumption can be applied to the interpretation of DSTI and LTV ratios, where we also observe a gradual decrease in indebtedness over time. The near-zero minimums in current lending conditions reflect the scenario where, towards the end of the repayment term, only a minimal debt amount remains outstanding for the borrower.

The variable representing education levels averages 3.2 with a median of 3: most borrowers have attained either a secondary education with graduation or a university degree. The current interest rate averages 2.4% with a standard deviation of 0.6078, which proposes relatively stable rates offered to borrowers.

Demographic data shows that the average number of debtors is 1.6 per loan. Furthermore, the average current age of applicants is 39.6 years, with a wide range between 18 to 78 years. The average current age of the loans in our dataset is 3.3 years, with the longest only being 12.3 years, which reminds us that we do not have loans that have undergone the entire standard repayment cycle.

Macroeconomic indicators are also considered, the GDP grows on average by 2.0%, and there's an average 3.9% rise in unemployment. The variations in GDP, as demonstrated by a standard deviation of 3.5485, imply economic fluctuations throughout the observation window, which may influence loan performance metrics.

4 Methodology and Models

In this chapter, we discuss the methodology adopted in this thesis. Recognizing the complexity of the data under study, we focus on multiple statistical techniques. These include logistic regression, ordered logistic regression, rare events methods, and model testing. The aim is to ensure a comprehensive and reliable analysis that provides insightful conclusions. The following section presents the models that are employed. Besides the full models, the work also examines data segmented by applicant incomes, regions of origin, or DSTI and LTV levels. The last part summarizes the hypotheses.

4.1 Methods

4.1.1 Logistic Regression

Based on the nature of the dependent variable, we focus on binary response models, particularly logistic regression also called logit model. Logit model is a statistical method used to predict the probability of a binary response based on one or more independent variables.

The logit model is more widely used than the probit model. The paper by Jin et al. (2005) favours the logit for several reasons. Compared to probit models, logit models are preferred for their ease of interpretation, as the coefficients reflect alterations in the odds of the dependent variable instead of probability changes. Further, logit models are computationally less demanding and simpler to estimate than probit. Finally, the study finds that logit model handles rare event data more effectively.

An alternative to the logit/probit model is the linear probability model (LPM). However, unlike binary response models, the LPM may produce estimated probabilities that fall outside the range of the binary variable .

Both Wooldridge (2012) and Greene (2018) outline the following methodology.

The logit model is built around a logistic function that models the probability that the dependent variable equals a case (in our thesis, denoted by 1). This probability is expressed as:

$$P(y = 1 \mid x) = G(\beta x) = p(x),$$

where $P(y = 1 \mid x)$ is the conditional probability that the dependent variable y equals 1 given the independent variables x, β represents the vector of coefficients (including the intercept), and G is the logistic function. The logistic function G is defined as:

$$G(z) = \Lambda(z) = \frac{exp(z)}{1 + exp(z)}.$$

This function is the cumulative distribution function of the logistic distribution, which outputs values between 0 and 1 for any input value z, which is the linear combination of the independent variables and the corresponding coefficients:

$$G(\beta_0 + X\beta) = G(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k).$$

The coefficients can be interpreted in terms of odds ratios. An odds ratio greater than 1 indicates the event becomes more likely with each unit increase of the variable; an odds ratio smaller than 1 indicates the opposite. Packages in Stata and R, statistical software for estimating the models, use log-odds metrics. A positive coefficient of log-odds implies an increase in event probability with an increase in the variable's value, in case other variables remain unchanged. In contrast, a negative coefficient indicates a decrease in event probability with an increase in the variable's value. To obtain log-odds, we divide the probability of the event by the probability of the event not occurring:

$$\text{Log-odds} = ln(\frac{P(y=1 \mid x)}{1-P(y=1 \mid x)}.$$

Log-odds coefficients require exponential transformation to be converted to odds ratios:

$$\text{Odds Ratio} = e^{(coefficient)}.$$

4.1.2 Ordered Logistic Regression

As one of the robustness checks, we employ ordered logistic regression. The ordered logit model is designed to predict outcomes that have a natural ordering; in our case, CNB's ratings from *standard*, through *watch*, to *non-performing*. The cumulative logit function is utilized to model the cumulative probabilities of each category up to a certain threshold. The ordered logistic regression equation can be represented as:

$$\text{logit}(P(Y \leq k \mid X)) = \alpha_k + \beta_1 X_1 + ... + \beta_k X_k,$$

where $P(Y \leq k \mid X)$ denotes the cumulative probability of the outcome being less than or equal to category k; α_k represents the threshold parameter for category k; and $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients associated with predictor variables $X_1, X_2, \ldots, X_k$.

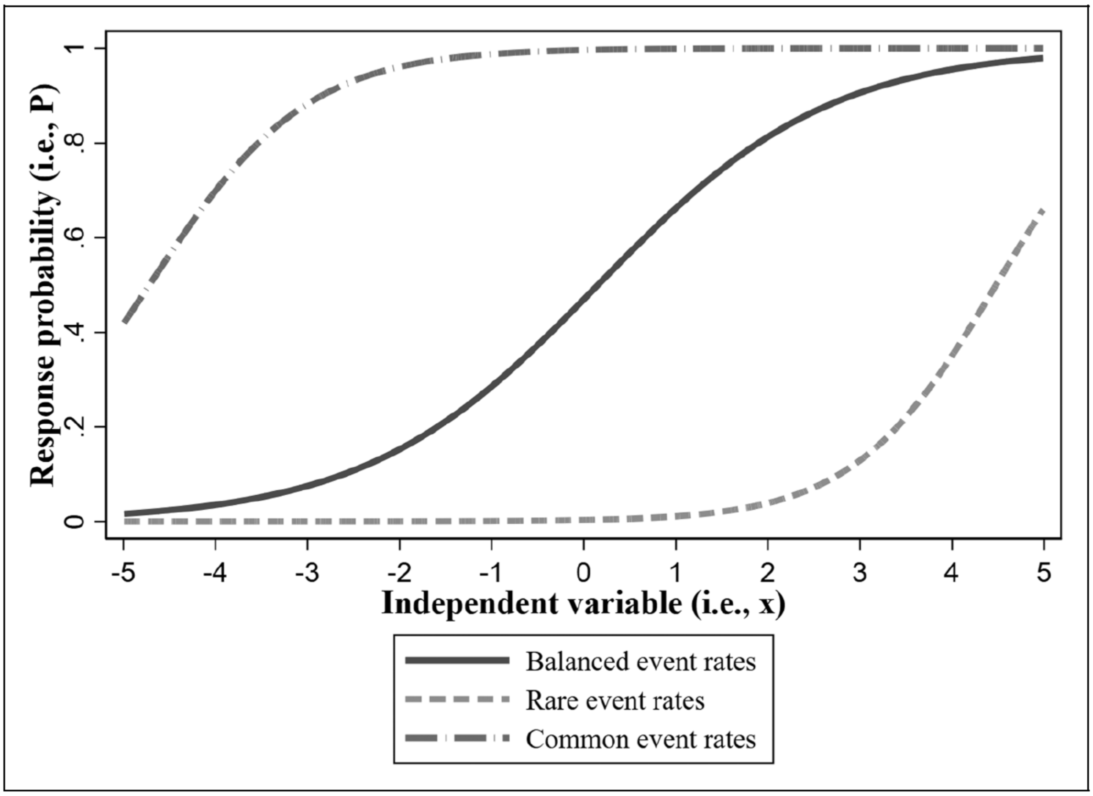
4.1.3 Penalized Logistic Regression

Due to the fact that our dependent variable is not equally distributed, we also focus on the rare event logistic models. One way to address rare events is through employing penalized models: Logistic regression with penalization may help in managing imbalanced data by penalizing the model for overfitting the majority class. A paper by Woo et al. (2022), which focuses on modelling binary dependent variables with rare or common events rate,[6] explains the problem on the response

[6] According to the paper, extreme rare event rates occur when the frequency of 1 is below 5%, while extreme common event rates are noted when the frequency of 1 is above 95%.

probability graphs of three cases (see Figure 12).

Figure 12: Response Probability Graphs



Source: Woo et al. (2022)

In models with balanced event rates, the fitted regression model appears as an S-shape curve, with response probabilities growing from 0 to 1 as the x value rises. For rare event rates (and analogically for common event rates), only a few event occurrences are observed, primarily when x values are very high. Therefore, response probabilities remain at or near 0 and escalate only at the high values of x. Such pronounced changes in response probabilities at either extreme of x, while remaining largely constant across its span, introduce inflation biases in logit and probit models.

We use two main alternative methods to tackle the issues encountered during the analysis of binary dependent variables under rare or common event rates with traditional binary response models: Firth logit (Firth, 1993) and Rare Event (RE) logit models (King and Zeng, 2001). Our thesis uses both methods as another robustness check. While Firth logit model employs a penalized MLE approach to correct the bias, RE logit model uses a weighted MLE approach.

Nevertheless, Woo et al. (2022) indicate that the bias decreases as the dataset's number of observations increases. They advise employing these alternative models when the sample size is under 500. For larger datasets, logit and probit models are appropriate.

4.2 Specification of the Model

Our analysis explores the impact of retail-lending parameters, particularly DSTI and LTV limits, on credit risk (repayment) using individual-level data. The base logit model takes the following form:

$$\begin{gathered}
\mathit{Delinquent\ Loan}_{i,t+1} = \alpha_i + \beta_1 \times \mathit{ODSTI}_i + \beta_2 \times \mathit{OLTV}_i + \beta_3 \times \mathit{Education\ Level}_i + \\
+ \beta_4 \times \mathit{Interest\ Rate}_{i,t} + \beta_5 \times \mathit{Number\ of\ Debtors}_i + \beta_6 \times \mathit{Age\ of\ Main\ Borrower}_{i,t} + \\
+ \beta_7 \times \mathit{Loan\ Age}_{i,t} + \beta_8 \times \mathit{GDP}_{i,t-3} + \beta_9 \times \mathit{Type\ of\ Profession}_i + \\
+ \beta_{10} \times \mathit{Marital\ Status}_i + \varepsilon_i,
\end{gathered}$$

where i denotes the client and t is the time index. The dependent variable is the delinquent loans delayed by one month to align with the 30-day past due criterion for categorization as 1. For the non-performing loans, a three-month lag is used, which reflects the 90-day overdue threshold. The further models for robustness check explore the delinquent and non-performing loans adjusted for an annual lag, as well as CNB rating, which is an ordered dependent variable.

Retail-lending parameters at the time of loan origination are represented by *ODSTI* and *OLTV*; in subsequent models, we test variable *ODTI*, along with various combinations of these metrics. We also assess the effect of the *current DTI, DSTI*, and *LTV*; however, it's important to note that these measures are estimated and should be considered when interpreting the results. Other variables that describe the borrowers' profile are *Level of Education*, *Number of Debtors*, *Age of Main Borrower*, *Marital Status*, and *Type of Profession. Level of Education* refers to the highest level of education the main borrower has attained by the time they apply for the mortgage. *Number of Debtors* counts the total number of applicants for the mortgage. *Age of Main Borrower* captures the current age of the main borrower. Furthermore, *Marital Status* indicates whether the main debtor was single, married, or otherwise at the time of the mortgage application. Lastly, *Type of Profession* details the profession of the main borrower at origination. Loan-based characteristics include *Interest Rate*, which is the current interest rate of the loan, and *Loan Age*, the age of the loan in months.

Finally, the macro-financial variable in our baseline equation is *GDP*, shifted by three months in comparison to the other of the variables. This shift assumes that changes in GDP do not immediately impact clients' behaviour but manifest with a delay. Apart from this three-month lag, we also examine the effects of varying lag durations.[7] In later models, we test the impact of unemployment.

4.3 Validation Techniques

One of the main reasons for the integration of multiple methods and a wide range of models is that it gives us a solid basis for performing robustness checks. This approach allows us to test the stability and consistency of our model results. Further, we aim to ensure that our results are not biased by specific model configurations.

[7] To summarize the lags, we offer a specific instance involving client Abc, whose delinquency dates to September, with other attributes recorded in August or at the time the loan was issued, and the value of GDP from May:

$$
\begin{aligned}
&\textit{Delinquent Loan}_{Abc,Sep} = \alpha_{Abc} + \beta_1 \times \textit{ODSTI}_{Abc} + \beta_2 \times \textit{OLTV}_{Abc} + \beta_3 \times \textit{Education Level}_{Abc} + \\
&+ + \beta_4 \times \textit{Interest Rate}_{Abc,Aug} + \beta_5 \times \textit{Number of Debtors}_{Abc} \beta_6 \times \textit{Age of Main Borrower}_{Abc,Aug} + \\
&+\beta_7 \times \textit{Loan Age}_{Abc,Aug} + \beta_8 \times \textit{GDP}_{Abc,May} + \beta_9 \times \textit{Type of Profession}_{Abc} + \\
&+\beta_{10} \times \textit{Marital Status}_{Abc} + \varepsilon_{Abc},
\end{aligned}
$$

Secondly, we employ the pseudo-R-squared by McFadden (1974), which measures the overall performance of the logit model. McFadden's R-squared is one such measure, calculated as $1 - \frac{L_{ur}}{L_o}$, where L_{ur} represents the log-likelihood of the model with predictors, and L_o is the log-likelihood of a baseline model with only an intercept. Because the log-likelihood for a binary response model is always negative, $|L_{ur}|$ is less than or equal to $|L_o|$, ensuring that the pseudo-R-squared lies between zero and one Wooldridge (2012).

Thirdly, we use the ROC (Receiver Operating Characteristic) analysis, which assesses the accuracy of the prediction. The outcome of the analysis is a ROC curve that plots the sensitivity (true positive proportion) against the specificity (false positive proportion) over different threshold probabilities. Further, the diagonal line representing random guess strategy, complements the graph. Finally, the model's effectiveness is measured by the integral area under the ROC curve (AUC). Higher AUC values correspond to higher model superiority in predicting true negatives as negatives and true positives as positives, i.e. a perfect model would have an AUC equal to 1 (Pearce and Ferrier, 2000).

Fourthly, we provide cross-validation. In cross-validation, we divide the data into k random subsets called folds (in our analysis, we use 10 folds). Each of these folds serves as a test set and the rest of the data form a training set. The primary objective is to ensure that the model evaluates the new data correctly and avoids overfitting. The performance is estimated as the average over all k-test sets (Berrar, 2018).

Finally, to deal with the fact that we work with large data and multiple explanatory variables, we conclude the discussion by several techniques that should improve the potential inaccuracy in standards errors and p-values. These approaches include panel and cluster-based logit models to handle variation both between and within clusters, cross-sectional analysis where the results are estimated using the dataset, in which every loan is applied only once, and the Bayesian model averaging (BMA), which is the most efficient selection approach for logistic regressions (Wang et al., 2004).

4.4 Hypotheses

By employing the logit models previously described, we aim to test hypotheses that investigate the relationship between retail-lending parameters and household delinquency, as well as their differential impacts across Czech regions. The hypotheses are as follows:

> **Hypothesis 1:** The level of DSTI has a significant effect on household delinquency.
>
> **Hypothesis 2:** The level of LTV has a significant effect on household delinquency.
>
> **Hypothesis 3:** The significance and size of the effects differ across Czech regions.

Beyond testing the base model across the entire dataset, we explore the effects of DSTI and LTV on subsets segmented by applicants' income and the type of collateral property. Additionally, we focus on the sensitivity to the economy. Finally, besides examining the DSTI, we also test DTI. The preference for DSTI over DTI is due to two reasons. Firstly, we consider DTI as an alternative to DSTI; therefore, we expect similar influence. Secondly, DSTI is more frequently employed both in practice and in the literature.

5 Results

This chapter outlines the outcomes of the constructed logit models. The models are segmented into various sections that examine the effects of individual borrower-based metrics across different groups. Thus, the analysis utilizes a wide scope of data. Additionally, the wide range of models also serves to verify the robustness of our findings. The chapter concludes with tests of the base model and an evaluation of the hypotheses.

5.1 Retail-Lending Parameters

5.1.1 Retail-Lending Parameters at Origination

Beyond the base model introduced in the previous chapter, we also examine the impact of borrower-based metrics at origination separately. The following Table 2 illustrates the regression outcomes for five models. The first column details the model centred on ODTI, followed by the second and third columns, which are dedicated to ODSTI and OLTV, respectively. The fourth column reveals the outcomes from our base model, in which we incorporate both ODSTI and OLTV alongside additional independent variables. The final column encompasses all retail-lending parameters at origination. For additional model variants and the graphical results of Model 4 (base model), please refer to the Appendix (Models 17 and Table 15).

The effects of ODTI, ODSTI, and OLTV align with general expectations. For instance, de Haan and Mastrogiacomo (2020) find that OLTV and ODSTI ratios are positively associated with the probability of non-performance. Moreover, Saha et al. (2022) show that a higher LTV positively affects the probability of default.

Table 2: Retail-Lending Parameters at Origination

Variables	*Model1* Delinquent Loan	*Model2*	*Model3*	*Model4*	*Model5*
ODTI	0.099***				0.008***
	(0.002)				(0.002)
ODSTI		0.045***		0.045***	0.045***
		(0.0004)		(0.0004)	(0.0004)
OLTV			0.011***	0.010***	0.010***
			(0.0002)	(0.0003)	(0.0003)
Education Level	-0.635***	-0.599***	-0.629***	-0.615***	-0.617***
	(0.0071)	(0.0071)	(0.0071)	(0.0071)	(0.0071)
Current Interest Rate	0.498***	0.462***	0.483***	0.461***	0.463***
	(0.0053)	(0.0054)	(0.0054)	(0.0054)	(0.0054)
Number of Debtors	-0.675***	-0.798***	-0.678***	-0.795***	-0.790***
	(0.0131)	(0.0129)	(0.0131)	(0.0129)	(0.0129)
Current Age of Main Borrower	-0.004***	-0.014***	-0.003***	-0.007***	-0.007***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0008)
Loan Age	0.0152***	0.0153***	0.0155***	0.0152***	0.0152***
	(0.000166)	(0.000165)	(0.000166)	(0.000166)	(0.000166)
GDP_{t-3}	-0.006***	-0.008***	-0.006***	-0.007***	-0.007***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
P.Blue Collar Worker	reference				
P.Director/Business Owner	0.421***	0.450***	0.607***	0.521***	0.511***
	(0.0175)	(0.0173)	(0.0174)	(0.0174)	(0.0177)
P.Employee	-0.0281*	0.0733***	0.0785***	0.122***	0.116***
	(0.0163)	(0.0163)	(0.0163)	(0.0163)	(0.0165)
P.Employee in Private Sector	-0.0962***	-0.111***	-0.0562***	-0.0936***	-0.0953***
	(0.0152)	(0.0152)	(0.0152)	(0.0152)	(0.0152)
P.Other	0.199***	0.220***	0.291***	0.271***	0.267***
	(0.0305)	(0.0305)	(0.0305)	(0.0305)	(0.0305)
P.Public Servant	-0.299***	-0.337***	-0.263***	-0.311***	-0.311***
	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.276***	0.276***	0.308***	0.275***	0.271***
	(0.0212)	(0.0212)	(0.0212)	(0.0212)	(0.0212)
MR.Married	reference				
MR.Other	-0.069***	-0.013	-0.074***	-0.041**	-0.042**
	(0.0168)	(0.0165)	(0.0167)	(0.0164)	(0.0164)
MR.Single	-0.073***	-0.026*	-0.079***	-0.060***	-0.062***
	(0.0155)	(0.0152)	(0.0155)	(0.0152)	(0.0152)
Constant	-5.207***	-5.901***	-5.524***	-6.808***	-6.847***
	(0.0498)	(0.0490)	(0.0526)	(0.0541)	(0.0552)
Pseudo-R^2	0.0620	0.0798	0.0620	0.0828	0.0828

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Examining the coefficients from our models, we find that ODTI's coefficient of 0.099 converts to an odds ratio of roughly 1.104, which indicates a 10.4% growth in the probability of payment delay for each unit increase of ODTI. For ODSTI, the coefficient is 0.045, which equates to an odds ratio of about 1.046, thereby increasing the delinquency probability by nearly 4.6% per unit rise. Lastly, the OLTV coefficient stands at 0.011, which gives an odds ratio close to 1.011; thus, a one-unit enlargement in OLTV is linked to an estimated 1.1% increase in the likelihood of delinquency. Therefore, of all the parameters, OLTV shows the smallest influence.

Our base model, referred to as Model 4, affirms prior findings: rising ODSTI and OLTV values lead to a higher likelihood of delinquency; the coefficient values remain nearly unchanged. The final fifth model presents all retail-lending parameters collectively. Despite the anticipated correlation due to the similarity between DTI and DSTI, the results here still align with those of the previous models.

Across all models, we observe a similar trend in variables tracking education, interest rate, and the number of debtors. These variables' effects are consistent with conventional expectations. While an increase in education level and a higher number of debtors decrease the probability of delinquency, a higher interest rate boosts this probability. We are reminded that the interest rate in our models reflects the current rate, i.e. it includes changes from either refinancing or refixation.

The effect of the applicant's current age on loan performance is not consistent across the literature. While de Haan and Mastrogiacomo (2020) suggest a negative correlation between the current applicant's age and non-performance, Brehanu and Fufa (2008) argue that relatively older borrowers are likely to be more responsible than their younger counterparts. Our models also display a negative correlation. The base model indicates that for each additional year of the applicant's age, the likelihood of delinquency is about 0.7%.

According to the loan age, there is a tendency for older mortgages to have a higher delinquency risk. However, we should be mindful that in this dataset, mortgage age is limited to a maximum of 12 years (the dataset spans from 2013 to 2023, with the oldest mortgages originating in 2011). For more definitive conclusions, it would be preferable to have data covering the entire lifespan of the mortgages.

Although the GDP growth coefficients are smaller relative to borrower characteristics, negative correlation supports the general expectation that an advancing economy reduces the chances of delinquency. In the base model, the odds are 0.993, which means that a 1% increase in GDP growth is associated with a 0.7% decrease in the probability of delinquency, assuming all other variables remain constant. This result suggests a modest but protective effect of GDP growth on reducing delinquency risks.

Finally, the regression outputs show that the coefficients for dummy variables representing profession and marital status maintain relative consistency across all models. For professions, directors or business owners have a higher likelihood of loan delinquency than blue-collar workers, who serve as the baseline group. Employees, particularly those in the private sector, tend to have a lower propensity for delinquency, as reflected by their negative coefficients. Also, public servants are less likely to fall into delinquency. Regarding marital status, individuals who are single or classified as other show increased chances of loan delinquency when contrasted with married individuals, the standard comparison group, as evidenced by the positive coefficients. Given the reliability of married debtors, we may suggest that if one partner struggles with repayments, the other is inclined to provide help.

In all models, standard errors are very small, and all variables seems to be significant. This outcome is probably caused by the fact that we work with large data and multiple explanatory variables. We discuss this topic further in section 5.6.

The last indicator we examined is the pseudo-R-squared, calculated using McFadden's formula. This metric is more appropriately used for comparing different specifications of the same model rather than as an absolute measure of fit. The R-squared values are relatively low, which may be attributed to the large dataset. Moreover, Smith and McKenna (2013) report that pseudo-R-squared typically provides lower estimates than OLS R-squared. When comparing the models, it appears that Model 4, our base model, and Model 5 perform best.

5.1.2 Current Retail-Lending Parameters

In this section, we turn to the question of the roles played by DTI, DSTI, and LTV, which evolve over time. These variables are mostly estimated; see Section 3 for details.

Across all models, in Table 3, a positive correlation persists between retail-lending parameters and delinquency. The effect of CDTI compared to ODTI has slightly decreased, with the probability of delinquency falling from 10.4% to 8.4% per unit increase. In contrast, the influence of CDSTI and CLTV on non-payment has significantly escalated. Generally, current retail-lending parameters decrease over time; therefore, any increase in these parameters for any reason will likely to lead to serious repayment issues for the client. This pattern is evident in Models 7 and 8, where parameters are observed individually, and in Model 9, where they are combined (see Tables 18 Appendix for more combinations).

The impact of other monitored variables remained almost unchanged. The pseudo-R-squared has slightly worsened compared to the first part, making these models appear less robust. This outcome confirms our assumption that parameters collected at the mortgage application stage are more accurate.

Table 3: Current Retail-Lending Parameters

Variables	*Model6* Delinquent Loan	*Model7*	*Model8*	*Model9*
CDTI	0.081***			
	(0.0030)			
CDSTI		0.913***		0.411***
		(0.0525)		(0.0537)
CLTV			2.112***	2.086***
			(0.0282)	(0.0284)
Education Level	-0.618***	-0.619***	-0.627***	-0.629***
	(0.0071)	(0.0071)	(0.0071)	(0.0071)
Current Interest Rate	0.496***	0.476***	0.508***	0.504***
	(0.0054)	(0.0054)	(0.0053)	(0.0054)
Number of Debtors	-0.665***	-0.670***	-0.682***	-0.677***
	(0.0132)	(0.0132)	(0.0131)	(0.0131)
Current Age of Main Borrower	-0.005***	-0.010***	0.004***	0.004***
	(0.0008)	(0.0007)	(0.0008)	(0.0008)
Loan Age	0.017***	0.017***	0.022***	0.022***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***	-0.007***	-0.004***	-0.005***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)
P.Blue Collar Worker	reference			
P.Director/Business Owner	0.469***	0.469***	0.586***	0.559***
	(0.0174)	(0.0176)	(0.0173)	(0.0177)
P.Employee	0.018	0.016	0.050***	0.048***
	(0.0163)	(0.0163)	(0.0163)	(0.0163)
P.Employee in Private Sector	-0.086***	-0.087***	-0.070***	-0.075***
	(0.0151)	(0.0152)	(0.0152)	(0.0152)
P.Other	0.230***	0.215***	0.295***	0.284***
	(0.0305)	(0.0305)	(0.0305)	(0.0305)
P.Public Servant	-0.293***	-0.296***	-0.273***	-0.276***
	(0.0216)	(0.0216)	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.304***	0.293***	0.292***	0.286***
	(0.0212)	(0.0212)	(0.0212)	(0.0212)
MR.Married	reference			
MR.Other	-0.060***	-0.054***	-0.080***	-0.081***
	(0.0168)	(0.0168)	(0.0167)	(0.0167)
MR.Single	-0.057***	-0.055***	-0.068***	-0.071***
	(0.0155)	(0.0155)	(0.0154)	(0.0154)
Constant	-5.055***	-4.741***	-6.455***	-6.518***
	(0.0515)	(0.0489)	(0.0545)	(0.0551)
Pseudo-R^2	0.0597	0.0590	0.0687	0.0688

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

5.1.3 Combination of Origination and Current Retail-Lending Parameters

After testing the parameters independently, we now examine both variables together (see Table 4). While Model 10 reveals a significant positive correlation for ODTI, it does not for CDTI. Meanwhile, Models 11 and 12 indicate significant correlations for the original and current lending parameters, albeit with opposite signs. Comparing the outcomes from the previous two sections, where each parameter was tested separately, we believe there is a mutual correlation between the original and current parameters, which biases the results.

Models 13 and 14 advance our analysis by adding a new variable to the DSTI and LTV combination (a variable that represents the product of these parameters). This addition aims to examine the interaction between the two factors. While the variable does not achieve significance for parameters at origination, it becomes significant in subsequent models, where we observe a negative correlation. The result indicates that higher CDSTI and CLTV values might decrease delinquency likelihood. This finding aligns with Dietsch and Welter-Nicol's (2014) research, which suggests that such clients with elevated parameters typically have higher incomes and a lower default risk due to less financial constraint than those at the threshold levels of LTV and DSTI. We explore this relationship in more detail in the following subsection.

5.1.4 Retail-Lending Parameters at Origination by Groups

For a closer look at the relationships between ODSTI and OLTV, we divide the dataset into four groups based on median values: Group 1 consists of loans where both ODSTI and OLTV are below the median, Group 2 contains loans with ODSTI below the median but OLTV above it, and so on (see Figure 13 below). Generally, we expected that ODSTI and OLTV in Group 1 exhibit small or no influence on delinquency, while parameters in Group show the greatest positive

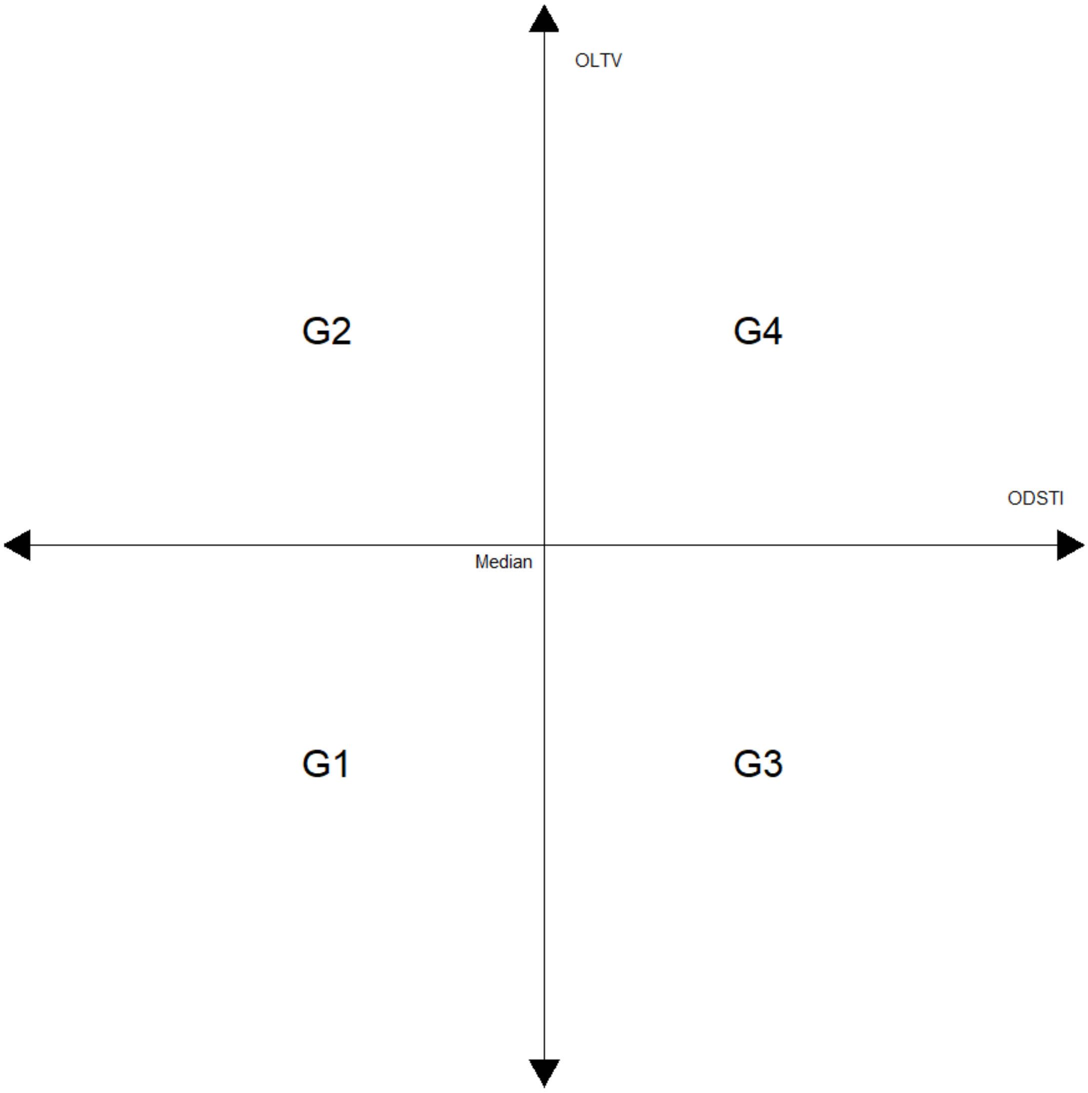
Table 4: Combination of Origination and Current Retail-Lending Parameters

Variables	*Model10* Delinquent Loan	*Model11*	*Model12*	*Model13*	*Model14*
ODTI	0.099***				
	(0.0026)				
CDTI	-0.0004				
	(0.0036)				
ODSTI		0.046***		0.046***	
		(0.0004)		(0.0013)	
CDSTI		-0.410***			4.483***
		(0.052)			(0.116)
OLTV			-0.001***	0.011***	
			(0.0003)	(0.0008)	
CLTV			2.219***		3.856***
			(0.0371)		(0.0544)
ODSTI * OLTV				-1.51e-05	
				(1.79e-05)	
CDSTI * CLTV					-8.434***
					(0.221)
Education Level	-0.635***	-0.596***	-0.625***	-0.615***	-0.626***
	(0.0071)	(0.0071)	(0.0071)	(0.0071)	(0.0071)
Current Interest Rate	0.498***	0.466***	0.509***	0.461***	0.510***
	(0.0054)	(0.0054)	(0.0053)	(0.0054)	(0.0054)
Number of Debtors	-0.675***	-0.810***	-0.682***	-0.795***	-0.677***
	(0.0131)	(0.0130)	(0.0131)	(0.0129)	(0.0131)
Current Age of Main Borrower	-0.004***	-0.014***	0.004***	-0.007***	0.005***
	(0.0008)	(0.0007)	(0.0008)	(0.0007)	(0.0007)
Loan Age	0.015***	0.015***	0.022***	0.015***	0.023***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.006***	-0.007***	-0.004***	-0.007***	-0.005***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
P.Blue Collar Worker	reference				
P.Director/Business Owner	0.421***	0.478***	0.578***	0.521***	0.525***
	(0.0175)	(0.0177)	(0.0174)	(0.0174)	(0.0177)
P.Employee	-0.028*	0.078***	0.044***	0.123***	0.049***
	(0.0163)	(0.0163)	(0.0163)	(0.0163)	(0.0162)
P.Employee in Private Sector	-0.096***	-0.106***	-0.073***	-0.094***	-0.074***
	(0.0152)	(0.0152)	(0.0152)	(0.0152)	(0.0152)
P.Other	0.199***	0.233***	0.290***	0.271***	0.273***
	(0.0305)	(0.0305)	(0.0305)	(0.0305)	(0.0305)
P.Public Servant	-0.299***	-0.337***	-0.276***	-0.311***	-0.268***
	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.276***	0.285***	0.291***	0.275***	0.276***
	(0.0212)	(0.0212)	(0.0212)	(0.0212)	(0.0212)
MR.Married	reference				
MR.Other	-0.069***	-0.010	-0.079***	-0.040**	-0.088***
	(0.0168)	(0.0164)	(0.0167)	(0.0164)	(0.0167)
MR.Single	-0.073***	-0.020	-0.064***	-0.059***	-0.078***
	(0.0155)	(0.0152)	(0.0154)	(0.0152)	(0.0154)
Constant	-5.205***	-5.817***	-6.421***	-6.851***	-7.410***
	(0.0515)	(0.0501)	(0.0551)	(0.0742)	(0.0598)
Pseudo-R^2	0.0620	0.0800	0.0688	0.0828	0.0713

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

coefficient for both ODSTI and OLTV.

Figure 13: Division of the Dataset into Four Groups



This assumption is mostly confirmed (find results in Table 5). In Group 4, ODSTI has the highest impact, where a 1% increase in this parameter is associated with a 7.6% increase in the probability of delinquency. The ODSTI coefficients for Group 2 and Group 3 are lower, indicating lesser risk for these groups. The coefficient in Group 1 is significant and negative, though very small, which suggests a negligible effect compared to other groups. Detailed analysis of Group 1 shows that the negative coefficient applies only to clients with higher incomes (see Tables 19 in Appendix for details). This observation aligns with the understanding that

Table 5: Retail-Lending Parameters at Origination by Groups

Variables	*Base* Delinquent Loan	*Group 1*	*Group 2*	*Group 3*	*Group 4*
ODSTI	0.045***	-0.004**	0.017***	0.050***	0.073***
	(0.0004)	(0.0015)	(0.0014)	(0.0007)	(0.0006)
OLTV	0.010***	0.026***	0.007***	0.006***	0.002**
	(0.0003)	(0.0007)	(0.0012)	(0.0004)	(0.0008)
Education Level	-0.615***	-0.937***	-1.060***	-0.739***	-0.610***
	(0.0071)	(0.0135)	(0.0119)	(0.0097)	(0.0083)
Current Interest Rate	0.461***	0.523***	0.576***	0.523***	0.605***
	(0.0054)	(0.0097)	(0.0096)	(0.0068)	(0.0062)
Number of Debtors	-0.795***	-0.393***	-0.874***	-0.542***	-1.164***
	(0.0129)	(0.0235)	(0.0227)	(0.0166)	(0.0150)
Current Age of Main Borrower	-0.007***	-0.025***	-0.029***	-0.015***	0.008***
	(0.0007)	(0.0014)	(0.0013)	(0.0009)	(0.0009)
Loan Age	0.015***	0.018***	0.012***	0.018***	0.013***
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***	0.011***	-0.011***	-0.007***	0.009***
	(0.0014)	(0.0027)	(0.0023)	(0.0019)	(0.0017)
P.Blue Collar Worker	reference				
P.Director/Business Owner	0.521***	0.744***	-0.333***	0.703***	0.652***
	(0.0174)	(0.0315)	(0.0488)	(0.0222)	(0.0206)
P.Employee	0.122***	0.029	-0.130***	0.167***	0.269***
	(0.0163)	(0.0325)	(0.0262)	(0.0251)	(0.0201)
P.Employee in Private Sector	-0.094***	0.181***	0.210***	0.433***	-0.162***
	(0.0152)	(0.0279)	(0.0217)	(0.0203)	(0.0181)
P.Other	0.271***	1.548***	-0.565***	0.038	-1.106***
	(0.0305)	(0.0392)	(0.0670)	(0.0447)	(0.0716)
P.Public Servant	-0.311***	-1.405***	0.447***	-0.093***	-0.584***
	(0.0216)	(0.0695)	(0.0282)	(0.0292)	(0.0270)
P.Senior and Medium Manager	0.275***	0.782***	-0.635***	0.492***	0.748***
	(0.0212)	(0.0387)	(0.0569)	(0.0300)	(0.0212)
MR.Married	reference				
MR.Other	-0.041**	0.961***	0.287***	-0.083***	-0.369***
	(0.0164)	(0.0306)	(0.0293)	(0.0226)	(0.0189)
MR.Single	-0.060***	0.657***	-0.003	0.426***	-0.379***
	(0.0152)	(0.0306)	(0.0279)	(0.0200)	(0.0172)
Loan Age	0.015***	0.018***	0.012***	0.018***	0.013***
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Constant	-6.808***	-5.689***	-3.298***	-6.779***	-7.242***
	(0.0541)	(0.103)	(0.135)	(0.0733)	(0.0927)
Pseudo-R^2	0.0828	0.131	0.111	0.110	0.117

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

higher-income clients generally have lower DSTI ratios and lower default risks. Regarding OLTV, we observe an opposite trend: while the lowest effect is observed in Group 4, the highest is in Group 1, contrary to expectations. If we further split Groups 1 and 4 by income, this phenomenon becomes clearer, showing that results are again influenced by income levels. Lower-income individuals in Group 1 experience a 3.7% increase in delinquency probability with a 1% increase in OLTV, while the higher-income group only shows a 1% increase. Further, in Group 4, the increase for lower-income individuals is 1.4%, but higher-income individuals show a negative trend, with a 1% OLTV increase leading to a 1% decrease in delinquency probability. This pattern confirms the study by Dietsch and Welter-Nicol (2014).

According to the higher pseudo-R-squared compared to the base model, splitting the dataset into these Groups also improved the performance of the models.

5.2 Region of Origination

Further, we explore the impact of ODSTI and OLTV across various Czech regions. We divided the dataset into 14 parts based on the applicants' residence address at origination. We assume that parameters' influence differs among regions, with a stronger impact expected in economically weaker regions that are considered more risky.

The model results confirm that sensitivity varies significantly across regions (Tables 6 and 7). Focusing on regions with the highest and lowest delinquency rates, Karlovy Vary stands out with the highest proportion of delinquent loans. It is the only region showing a negative correlation between OLTV and delinquency. This effect is very minor; a one percent increase in OLTV reduces the probability of delinquency by 0.3%, which assumes almost negligible influence of this parameter. Further examination of this region, using income-based segmentation, reveals that this negative correlation again appears among higher-income individuals (see

Table 6: Region of Origination (Part 1)

Variables	*Base* Delinquent Loan	*Prague*	*Central Bohemia*	*South Bohemia*	*Plzen*	*Karlovy Vary*	*Usti nad Labem*	*Liberec*
ODSTI	0.045***	0.046***	0.040***	0.047***	0.029***	0.037***	0.037***	0.033***
	(0.0004)	(0.0013)	(0.0012)	(0.0017)	(0.0020)	(0.0021)	(0.0017)	(0.0020)
OLTV	0.010***	0.003***	0.010***	0.012***	0.011***	-0.003*	0.017***	0.011***
	(0.0003)	(0.0007)	(0.0007)	(0.0011)	(0.0012)	(0.0014)	(0.0013)	(0.0013)
Education Level	-0.615***	-0.427***	-0.476***	-0.591***	-0.527***	-0.353***	-0.759***	-0.671***
	(0.0071)	(0.0225)	(0.0204)	(0.0317)	(0.0346)	(0.0405)	(0.0306)	(0.0348)
Current Interest Rate	0.461***	0.377***	0.346***	0.368***	0.457***	0.423***	0.504***	0.578***
	(0.0054)	(0.0166)	(0.0169)	(0.0270)	(0.0266)	(0.0304)	(0.0231)	(0.0235)
Number of Debtors	-0.795***	-0.944***	-0.460***	-0.990***	-0.732***	0.010	-0.910***	-1.492***
	(0.0129)	(0.0404)	(0.0340)	(0.0599)	(0.0605)	(0.0619)	(0.0571)	(0.0622)
Current Age of Main Borrower	-0.007***	0.0008	0.011***	0.020***	-0.014***	0.0007	-0.053***	-0.061***
	(0.0007)	(0.0022)	(0.0020)	(0.0032)	(0.0035)	(0.0039)	(0.0034)	(0.0039)
Loan Age	0.015***	0.020***	0.015***	0.014***	0.015***	0.008***	0.013***	0.018***
	(0.0002)	(0.0005)	(0.0005)	(0.0007)	(0.0008)	(0.0009)	(0.0007)	(0.0008)
GDP_{t-3}	-0.007***	0.009**	-0.011***	-0.004	-0.006	-0.017**	0.025***	0.006
	(0.0014)	(0.0041)	(0.0036)	(0.0060)	(0.0064)	(0.0075)	(0.0066)	(0.0070)
Type of Profession	yes	yes	yes	yes	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes	yes	yes	yes	yes
Constant	-6.808***	-7.134***	-8.496***	-7.792***	-6.404***	-7.103***	-4.686***	-2.794***
	(0.0541)	(0.176)	(0.153)	(0.243)	(0.252)	(0.292)	(0.244)	(0.258)
Pseudo-R^2	0.0828	0.0777	0.0544	0.0939	0.0634	0.0657	0.103	0.130

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Region of Origination (Part 2)

Variables	*Base* Delinquent Loan	*Hradec Kralove*	*Pardubice*	*Vysocina*	*South Moravia*	*Olomouc*	*Zlin*	*Moravia-Silesia*
ODSTI	0.045***	0.050***	0.054***	0.069***	0.058***	0.065***	0.048***	0.048***
	(0.0004)	(0.0022)	(0.0023)	(0.0022)	(0.0015)	(0.0018)	(0.0022)	(0.0012)
OLTV	0.010***	0.030***	0.019***	0.022***	0.019***	0.005***	0.008***	0.017***
	(0.0003)	(0.0016)	(0.0016)	(0.0016)	(0.0009)	(0.0012)	(0.0013)	(0.0008)
Education Level	-0.615***	-0.292***	-0.760***	-0.544***	-0.563***	-0.447***	-0.429***	-0.444***
	(0.0071)	(0.0402)	(0.0440)	(0.0426)	(0.0261)	(0.0348)	(0.0373)	(0.0231)
Current Interest Rate	0.461***	0.500***	0.709***	-0.0001	0.371***	0.274***	0.157***	0.498***
	(0.0054)	(0.0282)	(0.0277)	(0.0439)	(0.0204)	(0.0291)	(0.0362)	(0.0162)
Number of Debtors	-0.795***	-1.562***	-0.593***	-0.639***	-0.703***	-0.373***	-1.077***	-1.016***
	(0.0129)	(0.0739)	(0.0707)	(0.0637)	(0.0438)	(0.0490)	(0.0763)	(0.0420)
Current Age of Main Borrower	-0.007***	0.001	-0.001	-0.024***	-0.003	-0.028***	0.024***	0.002
	(0.0007)	(0.0040)	(0.0044)	(0.0044)	(0.0027)	(0.0037)	(0.0038)	(0.0023)
Loan Age	0.015***	0.013***	0.013***	0.014***	0.016***	0.018***	0.014***	0.014***
	(0.0002)	(0.0009)	(0.0010)	(0.0010)	(0.0006)	(0.0008)	(0.0009)	(0.0005)
GDP_{t-3}	-0.007***	-0.031***	0.043***	-0.040***	-0.023***	-0.022***	-0.0159**	-0.035***
	(0.0014)	(0.0069)	(0.0092)	(0.0069)	(0.0047)	(0.0060)	(0.0067)	(0.0040)
Type of Profession	yes	yes	yes	yes	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes	yes	yes	yes	yes
Constant	-6.808***	-8.695***	-9.247***	-7.563***	-8.088***	-7.904***	-7.821***	-7.642***
	(0.0541)	(0.303)	(0.324)	(0.320)	(0.198)	(0.254)	(0.299)	(0.172)
Pseudo-R^2	0.0828	0.114	0.107	0.0983	0.100	0.105	0.0737	0.107

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8). The impact of ODSTI in this region is consistent across the groups since there is about 3.8% increase in delinquency probability for each 1% increase in ODSTI.

In contrast, the Pardubice region has the smallest share of delinquent loans. Contrary to expectations that this region is less risky, the influence of ODSTI and OLTV is not diminished but is slightly higher than in the base model. This outcome indicates that delinquent loans, though few, are associated with higher ODSTI and OLTV when they do occur.

Regions with the most significant impact on parameter changes are Hradec Kralove and Vysocina. In Hradec Kralove, a one percent increase in OLTV raises the probability of delinquency by 3%, while Vysocina shows that a one percent increase in ODSTI boosts the delinquency probability by 9.9%. These regions demonstrate high sensitivity to financial metrics, which suggests banks should closely monitor DSTI and LTV values during client's evaluation.

Praha and Plzen are the most resilient regions regarding to loan repayments. In Praha, the impact of OLTV on the probability of delinquency drops to just 0.3%, which indicates that higher loans do not pose a significant risk. This development may be influenced by the fact that properties in Praha are the most expensive in the Czech Republic, thereby typically having higher LTV ratios, and by the greater financial capacity of its borrowers, who benefit from higher average incomes. Plzen exhibits the strongest resilience to changes in ODSTI, with a 1% increase in ODSTI correlating to a 3% increase in delinquency probability, which is still a substantial but manageable influence.

Please note that from this section, variables *Type of Profession* and *Marital Status* are included in the regressions, but the results are not displayed to keep the output clearer. More granular regional analyses are available upon request.

Table 8: Detail of the Karlovy Vary Region

Variables	*Base* Delinquent Loan	*Karlovy Vary*	*Income =< Median*	*Income > Median*
ODSTI	0.045***	0.037***	0.038***	0.039***
	(0.0004)	(0.0021)	(0.0030)	(0.0033)
OLTV	0.010***	-0.003*	0.006***	-0.011***
	(0.0003)	(0.0014)	(0.0019)	(0.0022)
Education Level	-0.615***	-0.353***	-0.647***	-0.016
	(0.0071)	(0.0405)	(0.0541)	(0.0637)
Current Interest Rate	0.461***	0.423***	0.441***	0.302***
	(0.0054)	(0.0304)	(0.0414)	(0.0481)
Number of Debtors	-0.795***	0.010	-0.500***	0.399***
	(0.0129)	(0.0619)	(0.104)	(0.0919)
Current Age of Main Borrower	-0.007***	0.0007	-0.030***	0.030***
	(0.0007)	(0.0039)	(0.0053)	(0.0062)
Loan Age	0.015***	0.008***	0.013***	0.002
	(0.0002)	(0.0009)	(0.0012)	(0.0016)
GDP_{t-3}	-0.007***	-0.017**	-0.020**	-0.014
	(0.0014)	(0.0075)	(0.0097)	(0.0119)
Type of Profession	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes
Constant	-6.808***	-7.103***	-4.798***	-9.351***
	(0.0541)	(0.292)	(0.388)	(0.489)
Pseudo-R^2	0.0828	0.0657	0.0838	0.0697

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.3 Initial Income

In this part of the analysis, we divide the dataset based on initial income into three subsets: the lower 25% as the low-income, the upper 25% as the high-income, and the remaining 50% as the middle-income group. We assume that the low-income group would have a poorer repayment performance. This theory is confirmed by the intercept, which is lowest in the high-income group and the highest in the low-income group. Similar observations were made by Quercia et al. (2012), who found that households with lower incomes had higher default rates and were less likely to prepay their loans. The OLTV's impact on loan repayment

Table 9: Initial Income

Variables	*Base* Delinquent Loan	*Low Income*	*Middle Income*	*High Income*
ODSTI	0.045***	0.024***	0.047***	0.061***
	(0.0004)	(0.0006)	(0.0005)	(0.0008)
OLTV	0.010***	0.009***	0.010***	0.009***
	(0.0003)	(0.0003)	(0.0003)	(0.0005)
Education Level	-0.615***	-0.825***	-0.841***	-0.592***
	(0.0071)	(0.0108)	(0.0094)	(0.0149)
Current Interest Rate	0.461***	0.482***	0.492***	0.535***
	(0.0054)	(0.0084)	(0.0071)	(0.0089)
Number of Debtors	-0.795***	-0.449***	-0.703***	-1.567***
	(0.0129)	(0.0207)	(0.0157)	(0.0242)
Current Age of Main Borrower	-0.007***	-0.025***	-0.027***	0.030***
	(0.0007)	(0.0011)	(0.0010)	(0.0015)
Loan Age	0.015***	0.013***	0.018***	0.019***
	(0.0002)	(0.0002)	(0.0002)	(0.0003)
GDP_{t-3}	-0.007***	-0.004*	-0.001	-0.001
	(0.0014)	(0.0020)	(0.0018)	(0.0028)
Type of Profession	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes
Constant	-6.808***	-4.966***	-5.945***	-8.363***
	(0.0541)	(0.0789)	(0.0748)	(0.171)
Pseudo-R^2	0.0828	0.0673	0.104	0.194

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

is relatively stable across income groups, whereas the influence of ODSTI increases with higher income. This result may reflect banks' more cautious approach to applicants with lower incomes towards the risk of default due to insufficient earnings. It is noteworthy that the model dedicated to the high-income group outperforms all other models in the analysis, since it achieves the highest pseudo-R-squared value.

Details of models for each group, with separately tested parameters, can be found in Appendix - Tables 9.

5.4 Property Type

We further divide the dataset into three categories to test robustness, based on the type of secured property: House, Flat, and Other. Results across these groups remain consistent, except for the Other category, where the influence of OLTV changes from positive to negative correlation. This category, which includes mortgages for lands, garages, or cottages, represents only 4% of the dataset. Thus, to draw more definitive conclusions, a larger sample would be necessary. However, it appears that clients in this category behave differently, probably because these loans are not primarily for residential properties but possibly for investment purposes. Compared to the other two groups, Other properties have a higher de-

Table 10: Property Type

Variables	*Base* Delinquent Loan	*House*	*Flat*	*Other*
ODSTI	0.045***	0.047***	0.040***	0.051***
	(0.0004)	(0.0006)	(0.0006)	(0.0015)
OLTV	0.010***	0.017***	0.008***	-0.006***
	(0.0003)	(0.0004)	(0.0004)	(0.0010)
Education Level	-0.615***	-0.578***	-0.677***	-0.260***
	(0.0071)	(0.0103)	(0.0106)	(0.0280)
Current Interest Rate	0.461***	0.433***	0.460***	0.477***
	(0.0054)	(0.0084)	(0.0082)	(0.0163)
Number of Debtors	-0.795***	-0.874***	-0.719***	-0.727***
	(0.0129)	(0.0177)	(0.0203)	(0.0469)
Current Age of Main Borrower	-0.007***	-0.003***	-0.013***	-0.019***
	(0.0007)	(0.0011)	(0.0011)	(0.0028)
Loan Age	0.015***	0.016***	0.014***	0.021***
	(0.0002)	(0.0003)	(0.0003)	(0.0007)
GDP_{t-3}	-0.007***	-0.008***	-0.005**	-0.011**
	(0.0014)	(0.0019)	(0.0021)	(0.005)
Type of Profession	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes
Constant	-6.808***	-7.464***	-6.020***	-7.527***
	(0.0541)	(0.0769)	(0.0841)	(0.216)
Pseudo-R^2	0.0828	0.0879	0.0737	0.174

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

linquency rate; we suggest that these loans might be prioritized lower for repayment during financial struggles. More detailed models are available upon request.

5.5 Economic Sensitivity

In the last section, we introduce models focused on economic parameters. While the base model used GDP growth shifted by three months, the model in the middle column uses GDP growth shifted by six months. The influence and significance of the variable remain unchanged. Still with the longer time-shift, we observe that

Table 11: Economic Sensitivity

Variables	*Base Model* Delinquent Loan	*GDP*	*Unemployment*
ODSTI	0.045***	0.045***	0.045***
	(0.0004)	(0.0004)	(0.0004)
OLTV	0.010***	0.010***	0.010***
	(0.0003)	(0.0003)	(0.0003)
Education Level	-0.615***	-0.615***	-0.615***
	(0.0071)	(0.0071)	(0.0071)
Current Interest Rate	0.461***	0.465***	0.465***
	(0.0054)	(0.0054)	(0.0056)
Number of Debtors	-0.795***	-0.795***	-0.795***
	(0.0129)	(0.0129)	(0.0129)
Current Age of Main Borrower	-0.007***	-0.008***	-0.008***
	(0.0007)	(0.0007)	(0.0007)
Loan Age	0.015***	0.015***	0.015***
	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***		
	(0.0014)		
GDP_{t-6}		-0.023***	
		(0.0013)	
$Unemployment_{t-3}$			-0.012**
			(0.0052)
Type of Profession	yes	yes	yes
Marital Status	yes	yes	yes
Constant	-6.808***	-6.770***	-6.787***
	(0.0541)	(0.0540)	(0.0579)
Pseudo-R^2	0.0828	0.0833	0.0827

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

a 1% increase in GDP results in a 2.2% decrease in the probability of delinquency, compared to just a 0.7% decrease in the base model. The pseudo-R-squared for this model has also slightly improved. The second economic parameter we employ is the unemployment rate, which, similar to GDP growth, uses a time lag. Contrary to expectations, the outcome suggests that increasing employment might marginally reduce delinquency risk. For a better understanding, further testing of this variable would be advisable, i.e. we might employ another time shift. Finally, the influence of ODSTI and OLTV remains consistent with our hypotheses in these models.

5.6 Validation of the Base Model

As we finish the regression analyses, we focus on validation of the base model. Since the dependent variable is a rare event, we employ two specific approaches to control for potential bias. The first method is the Firth model. Consistency in p-values across both the Firth and base models indicates that variable significance remains unaffected by the penalization applied in the Firth model. Furthermore, after rounding, the standard errors also yield identical results. Finally, the penalized log-likelihood in the Firth model is less negative than the standard log-likelihood, which means that the Firth model, with its bias correction, slightly outperforms the unpenalized model, although the improvement is modest.

The second model we use to control the rare event-dependent variable is the RE model. The RElogit package that we use in Stata does not allow the employment of qualitative nominal variables, which is why the models are shortened. Even here, the coefficients and p-value remain the same; only the standard errors differ slightly. However, these discrepancies are minimal and do not seem to affect the overall interpretation of the model. Overall, we can state that rare events do not cause bias, and the results of the previous models are credible.

Table 12: Firth Model

	Logit	*Firth logit*
Variables	Delinquent Loan	
ODSTI	0.045***	0.045***
	(0.0004)	(0.0004)
OLTV	0.010***	0.010***
	(0.0003)	(0.0003)
Education Level	-0.615***	-0.615***
	(0.0071)	(0.0071)
Current Interest Rate	0.461***	0.461***
	(0.0054)	(0.0054)
Number of Debtors	-0.795***	-0.795***
	(0.0129)	(0.0129)
Current Age of Main Borrower	-0.007***	-0.007***
	(0.0007)	(0.0007)
Loan Age	0.015***	0.015***
	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***	-0.007***
	(0.0014)	(0.0014)
P.Blue Collar Worker	reference	
P.Director/Business Owner	0.521***	0.521***
	(0.0174)	(0.0174)
P.Employee	0.122***	0.122***
	(0.0163)	(0.0163)
P.Employee in Private Sector	-0.0936***	-0.0936***
	(0.0152)	(0.0152)
P.Other	0.271***	0.271***
	(0.0305)	(0.0305)
P.Public Servant	-0.311***	-0.311***
	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.275***	0.275***
	(0.0212)	(0.0212)
MR.Married	reference	
MR.Other	-0.041**	-0.041**
	(0.0164)	(0.0164)
MR.Single	-0.060***	-0.060***
	(0.0152)	(0.0152)
Constant	-6.808***	-6.808***
	(0.0541)	(0.0541)
(Penalized) Log likelihood	-269,989.68	-269,897.74

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: RE Model

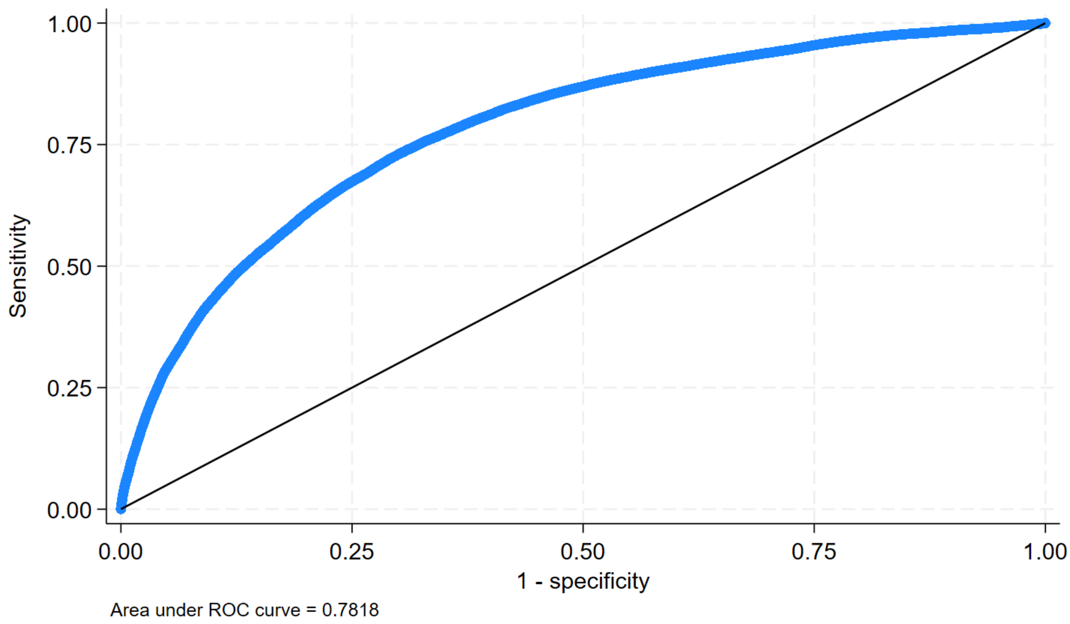
Variables	*Logit* Delinquent Loan	*RElogit*
ODSTI	0.045***	0.045***
	(0.0004)	(0.0004)
OLTV	0.009***	0.009***
	(0.0002)	(0.0002)
Education Level	-0.616***	-0.616***
	(0.0064)	(0.0065)
Current Interest Rate	0.469***	0.469***
	(0.0055)	(0.0048)
Number of Debtors	-0.747***	-0.747***
	(0.0098)	(0.0107)
Current Age of Main Borrower	-0.004***	-0.004***
	(0.0006)	(0.0006)
Loan Age	0.014***	0.014***
	(0.0002)	(0.0001)
GDP_{t-3}	-0.010***	-0.010***
	(0.0013)	(0.0014)
Constant	-6.862***	-6.862***
	(0.0445)	(0.0478)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Further testing of the base model includes ROC curve and cross-validation analysis. The AUC, which measures the overall performance, stands at 0.7818. According to Pearce and Ferrier (2000), values between 0.7 and 0.9 denote adequate discrimination; therefore, our base model indicates a good ability to distinguish between the two classes. In contrast, the cross-validation analysis reveals quite high sensitivity and low specificity. That means the model successfully predicts repayment but is less accurate in identifying actual delinquency. This result corresponds to the fact that the dependent variable is a rare event, and the dataset is unbalanced with numerous zeros and a small percentage of ones, which could skew the analysis.

To conclude the discussion about the validation of our model, we prepare several more models that are included in the Appendix: Tables 22, 23, 24, and 25. Since we work with large data and have many observations, the standard errors and

Figure 14: ROC Curve and AUC



p-values can be affected. Another issue we should keep in mind is that data from one mortgage can be dependent on another. To deal with this matter, we use several techniques.

Firstly, we employ panel and cluster-based logit models. According to Sohn and Kim (2007), random-effects panel logit models are utilized to handle variation both between and within clusters (in our case, mortgages). We prefer the random-effects over the fixed-effects model because our models include time-invariant variables, which the fixed-effects model exclude due to their lack of variation. Furthermore, by using cluster-based model, we confirm that there is no internal correlation in the base model. We cluster standard errors at the mortgage level, which leads to a better estimation of these errors (Jayatillake et al., 2011).

An additional model is provided on the dataset, in which every loan is applied only once. For selecting a specific observation from the given mortgage, we establish different approaches. In the appendix, we present the results of the dataset in which each mortgage is selected from the month with the highest average delinquency (30 months), and if the mortgage duration is shorter than 30 months, the median value of its lifespan is used. This strategy significantly reduces number of observations in the dataset and serves as cross-section analysis and another robustness check (Beck et al., 1998).

Finally, we present a model with explanatory variables selected by BMA. The BMA identifies all possible models and uses the posterior probabilities of these models to conduct all inferences and predictions. This approach resolves collinearity issues that can occur due to the large number of explanatory variables (Wang et al., 2004).

All methods used confirm that the core of the analysis - the magnitude and sign of the coefficients - remains the same, but they improve the standard errors and p-values.

5.7 Evaluation of Hypotheses

In the final part of this chapter, we will evaluate the hypotheses that we established at the beginning of the thesis.

Hypothesis 1: The level of DSTI has a significant effect on household delinquency.

The hypothesis was confirmed across all models and subsets. The models tested both DSTI at origination and current DSTI in various combinations. Both parameters showed a positive significant correlation with loan delinquency. Moreover, DTI, which can be considered as an alternative to DSTI, also showed a positive significant correlation with non-repayment.

Hypothesis 2: The level of LTV has a significant effect on household delinquency.

Based on the results, the level of LTV has a positive significant impact on mortgage delinquency. However, this effect is lower than that of DSTI. Furthermore, we found that for clients with high income and high LTV, this parameter has no effect, or shows a negative correlation. These conclusions are consistent with Dietsch and Welter-Nicol (2014) findings.

Hypothesis 3: The significance and size of the effects differ across Czech regions.

The results indicate that the effect of parameters is in line with previous conclusions. However, the size of effects differs across Czech regions, and in one region, we discovered the opposite correlation for LTV, which corresponds to our hypothesis.

6 Robustness Check

As mentioned in Chapter 3, we use also other dependent variables and subsets for robustness check.

Table 14: Dependent Variables

Variable	Description	Time Shift (in months)	Condition
Delinquent Loan (DelL1)	1 indicating mortgage payment is 30 and more days past due or in default; 0 otherwise	$y_{t+1} \sim x_t$	all loans stay for the whole time period
Delinquent Loan (DelL2)	1 indicating mortgage payment is 30 and more days past due or in default; 0 otherwise	$y_{t+1} \sim x_t$	after the first appearance of 1, the mortgage is discarded
Delinquent Loan (DelL_y)	1 indicating mortgage payment is 30 and more days past due or in default; 0 otherwise	$y_{t+12} \sim x_t$	
Non-performing Loan (NPL1)	1 indicating mortgage payment is 90 and more days past due or in default; 0 otherwise	$y_{t+3} \sim x_t$	
Non-performing Loan (NPL_y)	1 indicating mortgage payment is 90 and more days past due or in default; 0 otherwise	$y_{t+12} \sim x_t$	

Our base model, which was the core model through analysis, corresponds to the *DelL1* variable. We adjust this variable twice as part of the robustness check. Firstly, in the dataset with *DelL2* variable, we exclude all observations following the first instance of a 1 for a given mortgage. Secondly, we present an annual shift instead of a monthly shift. With an annual shift, we evaluate the effects of variables over a longer time period on delinquency. Another variable we used in this section corresponds to non-performance. For this variable, we used a 3-month shift, which covers mortgages that are 90 days past due or in other types of defaults, and again an annual shift.

Table 15: Robustness Check

Variables	DelL1	DelL2	DelL_y	NPL1	NPL_y
ODSTI	0.045***	0.050***	0.045***	0.045***	0.045***
	(0.0004)	(0.0017)	(0.0004)	(0.0004)	(0.0004)
OLTV	0.010***	0.006***	0.009***	0.011***	0.010***
	(0.0003)	(0.0010)	(0.0003)	(0.0003)	(0.0003)
Education Level	-0.615***	-0.482***	-0.598***	-0.609***	-0.592***
	(0.0071)	(0.0303)	(0.0073)	(0.0074)	(0.0076)
Current Interest Rate	0.461***	0.542***	0.694***	0.487***	0.668***
	(0.0054)	(0.0328)	(0.0077)	(0.0061)	(0.0080)
Number of Debtors	-0.795***	-0.775***	-0.799***	-0.784***	-0.789***
	(0.0129)	(0.0545)	(0.0132)	(0.0134)	(0.0136)
Current Age of Main Borrower	-0.007***	-0.006*	-0.008***	-0.003***	-0.005***
	(0.0007)	(0.0031)	(0.0008)	(0.0008)	(0.0008)
Loan Age	0.015***	-0.0008	0.014***	0.016***	0.014***
	(0.0002)	(0.0009)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***	-0.063***	-0.025***	-0.021***	-0.028***
	(0.0014)	(0.0050)	(0.0012)	(0.0013)	(0.0013)
Type of Profession	yes	yes	yes	yes	yes
Marital Status	yes	yes	yes	yes	yes
Constant	-6.808***	-9.125***	-6.982***	-7.165***	-7.225***
	(0.0541)	(0.240)	(0.0578)	(0.0567)	(0.0599)
Pseudo-R^2	0.0828	0.0446	0.0794	0.0827	0.0782

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The results of the models do not deviate significantly and confirm the robustness of the base model. Particularly, regarding ODSTI, a 1% increase affects the probability of loan delinquency by 4.6%; in these additional models, the effect varies between 4.6-5.1%. A one percentage point rise in OLTV has an effect on delinquency of 1% in the base model, in the second model this effect drops to 0.6%, and finally, in the last three it again fluctuates around 1%.

The last model in Table 16 is related to the ordered logit model. The dependent variable is the CNB Rating, which has a value from 1 to 3, from the best to the worst assessment of the loan. Cut1 represents the threshold between the ratings 1 and 2/3, and Cut2 is between 1/2 and 3; these cuts are part of the model's intercepts. The interpretation of the results is not as straightforward as with ordinary logit models; however, it holds still true that with the increase in ODSTI and OLTV, the probability that the loan will be in a worse category also increases.

Table 16: Ordered Logit Model

Variables	CNB
ODSTI	0.038***
	(0.0003)
OLTV	0.0005***
	(0.0002)
Education Level	-0.548***
	(0.0051)
Current Interest Rate	0.510***
	(0.0039)
Number of Debtors	-0.515***
	(0.0090)
Current Age of Main Borrower	0.007***
	(0.0005)
Loan Age	0.009***
	(0.0001)
GDP_{t-3}	-0.005***
	(0.0010)
Type of Profession	yes
Marital Status	yes
/cut1	6.304***
	(0.0381)
/cut2	7.094***
	(0.0383)
Pseudo-R^2	0.0679

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

After running additional regressions, we may conclude that our evaluation of the hypothesis still holds.

7 Conclusion

This paper contributes to the debate on how effective are macroprudential policy instruments, specifically DTI, DSTI and LTV, and what their effect is on loan delinquency and default. These retail-lending parameters are increasingly popular among the authorities, who use them to support financial stability and limit the creation of excessive risks related to the mortgage market. Caps for these parameters, which central banks and regulators set, are intended to prevent banks from providing more risky loans.

The thesis presents unique monthly loan-level data from around 250 thousand mortgages from July 2013 to July 2023, which was harvested from one Czech commercial bank. This data contains applicant DTI, DSTI and LTV at origination, age of an applicant, interest rate and others. We complete the dataset by estimating current DTI, DSTI and LTV. Furthermore, we include macroeconomic variables such as GDP growth and unemployment rate in the dataset.

To uncover the effects of retail-lending parameters, we use logit models. The dependent variable indicated loan delinquency, where 1 represents loans 30 days past due or in default, and 0 otherwise. In the base model, we monitor the effect of DSTI and LTV at origination and compare these relationships across different subsets. The results confirm our hypotheses: both DSTI and LTV have a significant effect on household delinquency. According to the base model, increase of one percent in ODSTI raises the probability of delinquency by nearly 4.6%. The impact of the OLTV increase of one percent is slightly smaller, boosting the probability of delinquency by 1.1%. Moreover, models with current DSTI and LTV further amplify this effect. Tests for DTI show 10.4% growth in the probability of payment delay for each unit of origination parameter and 8.4% growth for the current parameter. Nevertheless, in the area of current parameters, there is still room for further research, as our current parameters are only estimated.

The impact of LTV changes for people with high income and high LTV; for this group, an increase in OLTV by one percent leads to a 1% decrease in delinquency probability. We believe that this group of clients can sustain higher debt levels, while at the same time, they have sufficient financial resources to reduce the risk of non-payment. This non-linear relationship is also confirmed in the study by Dietsch and Welter-Nicol (2014).

The models confirm general assumptions for the effects of other explanatory variables. On one hand, a higher level of education and number of co-debtors on one loan reduces the risk. On the other hand, a higher interest rate, which leads to higher repayments, increases the risk of delinquency. The results are unclear for the current age of the main borrower, where the correlation is significantly negative but very small. Regarding the loan age, there is a tendency for older mortgages to have a higher delinquency risk; however, we should be mindful that most loans in the dataset did not undergo the entire standard repayment process. Finally, in times of economic growth, as indicated by GDP, the risk of non-payment is lower.

Since the loans are focused exclusively on the Czech sector, the dataset only includes loans to Czech residents and for properties within the Czech Republic. Our hypothesis that the significance and size of the effects differ across Czech regions is confirmed. The most significant change occurs in the Karlovy Vary region. Here, the positive correlation of OLTV changes to a negative correlation for clients with higher income. This development is in line with previous findings.

Because our dependent variable represents a rare event above the logit model, we also use the Firth and the RE logit models, as suggested by Woo et al. (2022). These two methods are tailored for rare events datasets, allowing us to validate the base model results. We confirm that the outcomes of these tests were consistent, and the results of the previous models are credible without bias.

For the robustness check, we create two more dependent variables: the first binary variable monitored non-performance and the second ordered variable corresponded to the CNB rating categories. The effects of ODSTI and OLTV remain largely unchanged, and we affirm that our evaluation of the hypotheses still holds.

To conclude, we evaluate retail-lending parameters as very important parameters to monitor when providing a mortgage. Setting the upper limit of these parameters can prevent the risk of non-payment and potential defaults. Nevertheless, we recommend allowing some flexibility so that it is possible to exceed the cap for certain portions of the loan. This option mainly concerns clients in the upper income and wealth classes, where the level of these limits, particularly in the case of LTV, does not appear to be a driver of delinquency.

Bibliography

Agbemava, E., Nyarko, I. K., Adade, T. C. and Bediako, A. K. (2016), 'Logistic regression analysis of predictors of loan defaults by customers of non-traditional banks in ghana', *European Scientific Journal* **12**.

BCBS (2016), 'Prudential treatment of problem assets - definitions of non-performing exposures and forbearance'.

Beck, N., Katz, J. N. and Tucker, R. (1998), 'Taking time seriously: Time-series-cross-section analysis with a binary dependent variable', *American Journal of Political Science* **42**(4), 1260–1288.

Berrar, D. (2018), Cross-validation. ISBN 978-0-12-809633-8.

BIS (2017), 'IFRS 9 and expected loss provisioning - executive summary'.

Brehanu, A. and Fufa, B. (2008), 'Repayment rate of loans from semi-formal financial institutions among small-scale farmers in ethiopia: Two-limit tobit analysis', *The Journal of Socio-Economics* **37**(6), 2221–2230.

Campbell, J. Y. and Cocco, J. F. (2015), 'A model of mortgage default', *Journal of Finance* **70**(4), 1495–1554.

CNB (2024), 'Requirements for LTV, DSTI and DTI limits'. Accessed: 2024-04-23.
URL: *https://www.cnb.cz/en/financial-stability/macroprudential-policy/requirements-for-ltv-dsti-and-dti-limits/*

de Haan, L. and Mastrogiacomo, M. (2020), 'Loan to value caps and government-backed mortgage insurance: Loan-level evidence from dutch residential mortgages', *De Economist* **168**(4), 453–473.

Derviz, A. and Raková, M. (2012), 'Parent influence on loan pricing by czech banks', *Prague Economic Papers* **21**(4), 434–449.

Dietsch, M. and Welter-Nicol, C. (2014), 'Do LTV and DSTI caps make banks more resilient?', *Rue de la Banque* (21).

Dzidzeviciute, L. (2012), 'Estimation of default probability for low default portfolios', *Ekonomika* **91**, 132–156.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D. and Hunt, R. (2010), 'What "triggers" mortgage default?', *The American Economic Review* **100**(2), 490–494.

Emekter, R., Tu, Y., Jirasakuldech, B. and Lu, M. (2015), 'Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending', *Applied Economics* **47**(1), 54–70.

ESRB (2013), 'Recommendation of the european systemic risk board of 4 april 2013 on intermediate objectives and instruments of macro-prudential policy (ESRB/2013/1)'.
URL: *https://www.esrb.europa.eu/pub/pdf/recommendations/2013/ESRB_2013_1.en.pdf*

ESRB (2021), 'A review of macroprudential policy in the EU in 2020.'.
URL: *https://data.europa.eu/doi/10.2866/77568*

Fio (2021), 'Mezitimní účetní závěrka k 30. červnu 2021'.
URL: *https://www.fio.cz/docs/cz/Fio_banka_MUZ_20210630.pdf*

Firth, D. (1993), 'Bias reduction of maximum likelihood estimates', *Biometrika* **80**(1), 27–38.

Froyland, E. and Larsen, K. (2002), 'How vulnerable are financial institutions to macroeconomic changes? an analysis based on stress testing', *Economic Bulletin* **Q3**(2), 92–98.

Gaudêncio, J., Mazany, A. and Schwarz, C. (2019), 'The impact of lending standards on default rates of residential real-estate loans', *ECB Occasional Paper Series* **220**.

Gerlach-Kristen, P. and Lyons, S. (2018), 'Determinants of mortgage arrears in europe: evidence from household microdata', *International Journal of Housing Policy* **18**(4), 545–567.

Geršl, A. and Jakubík, P. (2011), 'Relationship lending in emerging markets: Evidence from the czech republic', *Comparative Economic Studies* .

Geršl, A., Jakubík, P., Kowalczyk, D., Ongena, S. and Peydro, J.-L. (2015), 'Monetary conditions and banks' behaviour in the czech republic', *Open Economies Review* **26**(3), 407–445. Publisher: Springer.

Greene, W. (2018), 'Econometric analysis'. ISBN 978-0-13-446136-6.

Greve, C. and Hahnenstein, L. (2014), 'Stress testing the credit risk of mortgage loans: The relationship between portfolio-LGD and the loan-to-value distribution'.

Gross, M. and Población, J. (2017), 'Assessing the efficacy of borrower-based macroprudential policy using an integrated micro-macro model for european households', *Economic Modelling* **61**, 510–528.

Herrala, R. and Kauko, K. (2007), 'Household loan loss risk in finland - estimations and simulations with micro data', *SSRN Electronic Journal* .

Jayatillake, R., Sooriyarachchi, R. and Senarathna, L. (2011), 'Adjusting for a cluster effect in the logistic regression model: An illustration of theory and its application', *Journal of the National Science Foundation of Sri Lanka* **39**.

Jin, Y., Rejesus, R. and Little, B. (2005), 'Binary choice models for rare events data: a crop insurance fraud application', *Applied Economics* **37**(7), 841–848.

Jurča, P., Klacso, J., Tereanu, E., Forletta, M. and Gross, M. (2020), 'The effectiveness of borrower-based macroprudential measures: A quantitative analysis for slovakia', *IMF Working Papers* **20**(134).

Kelly, R. and O'Toole, C. (2018), 'Mortgage default, lending conditions and macroprudential policy: Loan-level evidence from UK buy-to-lets', *Journal of Financial Stability* **36**.

King, G. and Zeng, L. (2001), 'Logistic regression in rare events data'.

McFadden, D. (1974), 'Conditional logit analysis of qualitative choice behavior'.

Pearce, J. and Ferrier, S. (2000), 'Evaluating the predictive performance of habitat models developed using logistic regression', *Ecological Modelling* **133**, 225–245.

Penikas, H. (2020), 'Low default portfolios in basel II and basel III as a special case of significantly unbalanced classes in binary choice models', *Russian Journal of Money and Finance* **79**(2), 101–128.

Qi, M. and Yang, X. (2008), 'Loss given default of high loan-to-value residential mortgages'.

Quercia, R. G., Pennington-Cross, A. and Yue Tian, C. (2012), 'Mortgage default and prepayment risks among moderate- and low-income households', *Real Estate Economics* **40**, 159–198.

Rahman, M. A. (2013), 'Household characteristics and poverty: A logistic regression analysis', *The Journal of Developing Areas* **47**(1), 303–317.

Saha, A., Rooj, D. and Sengupta, R. (2022), 'Loan to value ratio and housing loan default – evidence from microdata in india', *International Journal of Emerging Markets* .

Smith, T. J. and McKenna, C. M. (2013), 'A comparison of logistic regression pseudo r^2 indices', *Multiple Linear Regression Viewpoints* **39**.

Sohn, S. Y. and Kim, H. S. (2007), 'Random effects logistic regression model for default prediction of technology credit guarantee fund', *European Journal of Operational Research* **183**(1), 472–478.

Wang, D., Zhang, W. and Bakhai, A. (2004), 'Comparison of bayesian model averaging and stepwise methods for model selection in logistic regression', *Statistics in Medicine* **23**(22), 3451–3467.

Woo, H.-S., Berns, J. and Solanelles, P. (2022), 'How rare is rare? how common is common? empirical issues associated with binary dependent variables with rare or common event rates', *Organizational Research Methods* **26**, 109442812210831.

Wooldridge, J. M. (2012), 'Introductory econometrics: A modern approach'.

List of Tables

List of Figures

Acronyms

AUC	Area Under the ROC Curve
BMA	Bayesian Model Averaging
CDSTI	Current Debt Service-to-Income
CDTI	Current Debt-to-Income
CLTV	Current Loan-to-Value
CNB	Czech National Bank
CRC	Central Credit Register
CZK	Czech Koruna
CZSO	Czech Statistical Office
DSTI	Debt Service-to-Income
DTI	Debt-to-Income
ECL	Expected Credit Loss
ESRB	European Systemic Risk Board
GDP	Gross Domestic Product
HPI	Housing Price Index
LGD	Loss Given Default
LPM	Linear Probability Model
LTI	Loan-to-Income
LTV	Loan-to-Value

ODSTI	Originating Debt Service-to-Income
ODTI	Originating Debt-to-Income
OLS	Ordinary Least Squares
OLTV	Originating Loan-to-Value
ORC	Original Rent Coverage
PD	Probability of Default
RE	Rare Event
ROC	Receiver Operating Characteristic

Appendix

Figure 15: Coefficients of the Base Model (Table 2)

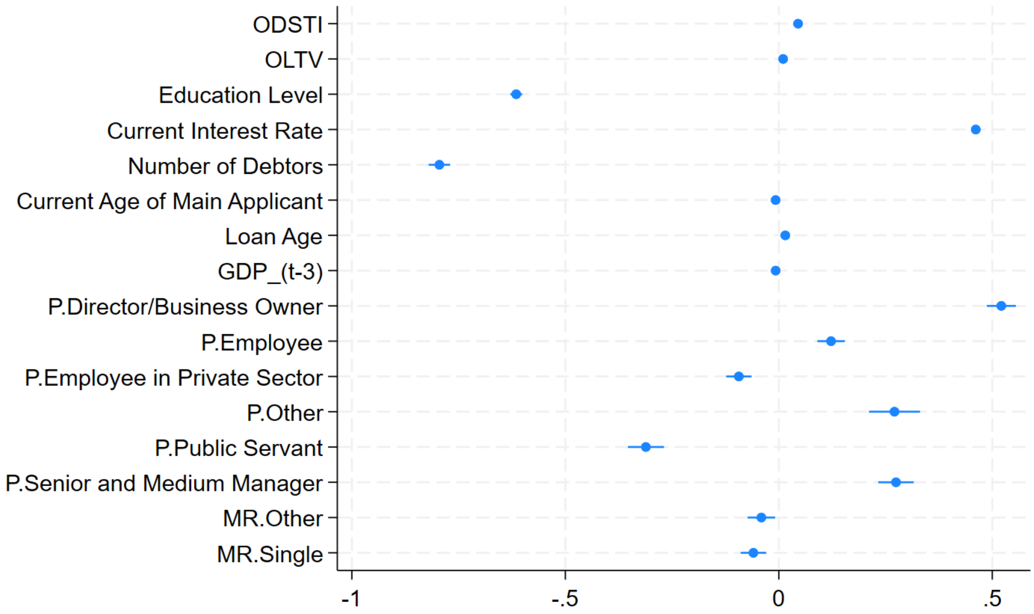


Table 17: Estimation of Other Model from Section 5.1.1 Retail-Lending Parameters at Origination

Variables	ModelA1 Delinquent Loan	ModelA2
ODTI	0.023***	0.085***
	(0.0023)	(0.0022)
ODSTI	0.044***	
	(0.0004)	
OLTV		0.001***
		(0.0003)
Education Level	-0.606***	-0.646***
	(0.0072)	(0.0071)
Current Interest Rate	0.466***	0.496***
	(0.0054)	(0.0054)
Number of Debtors	-0.786***	-0.673***
	(0.0130)	(0.0131)
Current Age of Main Borrower	-0.012***	0.001*
	(0.0007)	(0.0008)
Loan Age	0.015***	0.015***
	(0.0002)	(0.0002)
GDP_{t-3}	-0.008***	-0.006***
	(0.0014)	(0.0014)
P.Blue Collar Worker	reference	
P.Director/Business Owner	0.425***	0.504***
	(0.0175)	(0.0176)
P.Employee	0.057***	0.027*
	(0.0164)	(0.0164)
P.Employee in Private Sector	-0.115***	-0.078***
	(0.0152)	(0.0152)
P.Other	0.210***	0.252***
	(0.0305)	(0.0305)
P.Public Servant	-0.336***	-0.275***
	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.268***	0.280***
	(0.0212)	(0.0212)
MR.Married	reference	
MR.Other	-0.018	-0.085***
	(0.0165)	(0.0167)
MR.Single	-0.035**	-0.096***
	(0.0153)	(0.0154)
Constant	-6.035***	-5.975***
	(0.0509)	(0.0541)
Pseudo-R^2	0.0800	0.0644

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 18: Estimation of Other Model from Section 5.1.2 Current Retail-Lending Parameters

Variables	*ModelA3*	*ModelA4*
	Delinquent Loan	
CDTI	-0.005	-0.039***
	(0.0033)	(0.0044)
CDSTI		0.832***
		(0.0706)
CLTV	2.125***	2.171***
	(0.0297)	(0.0299)
Education Level	-0.626***	-0.630***
	(0.0071)	(0.0071)
Current Interest Rate	0.508***	0.495***
	(0.0054)	(0.0055)
Number of Debtors	-0.682***	-0.679***
	(0.0131)	(0.0131)
Current Age of Main Borrower	0.004***	0.003***
	(0.0008)	(0.0008)
Loan Age	0.022***	0.022***
	(0.0002)	(0.0002)
GDP_{t-3}	-0.004***	-0.005***
	(0.0014)	(0.0014)
P.Blue Collar Worker	reference	
P.Director/Business Owner	0.589***	0.563***
	(0.0175)	(0.0177)
P.Employee	0.051***	0.052***
	(0.0163)	(0.0163)
P.Employee in Private Sector	-0.070***	-0.076***
	(0.0152)	(0.0152)
P.Other	0.296***	0.280***
	(0.0305)	(0.0305)
P.Public Servant	-0.273***	-0.276***
	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.292***	0.280***
	(0.0212)	(0.0212)
MR.Married	reference	
MR.Other	-0.080***	-0.080***
	(0.0167)	(0.0167)
MR.Single	-0.067***	-0.071***
	(0.0154)	(0.0154)
Constant	-6.439***	-6.441***
	(0.0558)	(0.0557)
Pseudo-R^2	0.0687	0.0690

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 19: Estimation of Other Model from Section 5.1.4 Retail-Lending Parameters at Origination by Groups

Variables	*Group 1* Delinquent Loan	*ModelA5*	*ModelA6*	*ModelA7*	*ModelA8* Income =< Median	*ModelA9* Income > Median
ODTI		0.052***				
		(0.0045)				
ODSTI	-0.004**		0.003**		0.025***	-0.030***
	(0.0015)		(0.0015)		(0.0021)	(0.0022)
OLTV	0.026***			0.026***	0.037***	0.010***
	(0.0007)			(0.0006)	(0.0009)	(0.0010)
Education Level	-0.937***	-0.918***	-0.916***	-0.936***	-1.152***	-0.666***
	(0.0135)	(0.0136)	(0.0136)	(0.0135)	(0.0189)	(0.0215)
Current Interest Rate	0.523***	0.524***	0.524***	0.523***	0.600***	0.373***
	(0.0097)	(0.0096)	(0.0096)	(0.0097)	(0.0144)	(0.0142)
Number of Debtors	-0.393***	-0.397***	-0.386***	-0.393***	0.186***	-0.949***
	(0.0235)	(0.0235)	(0.0236)	(0.0236)	(0.0298)	(0.0382)
Current Age of Main Borrower	-0.025***	-0.029***	-0.032***	-0.025***	-0.034***	-0.020***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0019)	(0.0023)
Loan Age	0.018***	0.018***	0.018***	0.018***	0.016***	0.023***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0005)
GDP_{t-3}	0.011***	0.010***	0.010***	0.011***	0.008**	0.008**
	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.004)	(0.004)
P.Blue Collar Worker	reference					
P.Director/Business Owner	0.744***	0.702***	0.731***	0.742***	-0.177***	1.268***
	(0.0315)	(0.0315)	(0.0314)	(0.0315)	(0.0572)	(0.0519)
P.Employee	0.0288	0.0104	0.0203	0.0294	0.321***	-0.365***
	(0.0325)	(0.0326)	(0.0325)	(0.0326)	(0.0374)	(0.0671)
P.Employee in Private Sector	0.181***	0.140***	0.140***	0.182***	0.435***	-0.332***
	(0.0279)	(0.0279)	(0.0279)	(0.0279)	(0.0322)	(0.0570)
P.Other	1.548***	1.487***	1.523***	1.549***	2.068***	0.028
	(0.0392)	(0.0394)	(0.0391)	(0.0392)	(0.0445)	(0.100)
P.Public Servant	-1.405***	-1.435***	-1.429***	-1.405***	-1.661***	-1.128***
	(0.0695)	(0.0695)	(0.0695)	(0.0695)	(0.102)	(0.0992)
P.Senior and Medium Manager	0.782***	0.798***	0.800***	0.782***	-0.561***	0.859***
	(0.0387)	(0.0388)	(0.0388)	(0.0387)	(0.116)	(0.0573)
MR.Married	reference					
MR.Other	0.961***	0.978***	1.000***	0.963***	1.910***	-0.820***
	(0.0306)	(0.0307)	(0.0307)	(0.0306)	(0.0433)	(0.0653)
MR.Single	0.657***	0.678***	0.681***	0.657***	1.102***	0.466***
	(0.0306)	(0.0306)	(0.0307)	(0.0306)	(0.0448)	(0.0435)
Constant	-5.689***	-4.539***	-4.368***	-5.767***	-7.527***	-3.828***
	(0.103)	(0.0926)	(0.0970)	(0.0977)	(0.141)	(0.171)
Pseudo-R^2	0.131	0.121	0.120	0.131	0.163	0.162

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Variables	*Group 2* Delinquent Loan	*ModelA10*	*ModelA11*	*ModelA12*
ODTI		-0.010**		
		(0.0047)		
ODSTI	0.017***		0.017***	
	(0.0014)		(0.0014)	
OLTV	0.007***			0.007***
	(0.0012)			(0.0012)
Education Level	-1.060***	-1.058***	-1.061***	-1.058***
	(0.0119)	(0.0119)	(0.0119)	(0.0119)
Current Interest Rate	0.576***	0.588***	0.586***	0.580***
	(0.0096)	(0.0095)	(0.0095)	(0.0096)
Number of Debtors	-0.874***	-0.862***	-0.882***	-0.859***
	(0.0227)	(0.0227)	(0.0226)	(0.0226)
Current Age of Main Borrower	-0.029***	-0.029***	-0.030***	-0.028***
	(0.0013)	(0.0013)	(0.0013)	(0.0013)
Loan Age	0.012***	0.012***	0.012***	0.012***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
GDP_{t-3}	-0.011***	-0.010***	-0.010***	-0.010***
	(0.0023)	(0.0023)	(0.0023)	(0.0023)
P.Blue Collar Worker	reference			
P.Director/Business Owner	-0.333***	-0.365***	-0.367***	-0.337***
	(0.0488)	(0.0485)	(0.0485)	(0.0488)
P.Employee	-0.130***	-0.165***	-0.169***	-0.130***
	(0.0262)	(0.0254)	(0.0254)	(0.0261)
P.Employee in Private Sector	0.210***	0.212***	0.206***	0.213***
	(0.0217)	(0.0217)	(0.0217)	(0.0217)
P.Other	-0.565***	-0.588***	-0.582***	-0.571***
	(0.0670)	(0.0669)	(0.0669)	(0.0670)
P.Public Servant	0.447***	0.447***	0.435***	0.456***
	(0.0282)	(0.0282)	(0.0282)	(0.0282)
P.Senior and Medium Manager	-0.635***	-0.643***	-0.643***	-0.636***
	(0.0569)	(0.0569)	(0.0569)	(0.0569)
MR.Married	reference			
MR.Other	0.287***	0.302***	0.293***	0.297***
	(0.0293)	(0.0293)	(0.0293)	(0.0293)
MR.Single	-0.003	0.013	0.001	0.008
	(0.0279)	(0.0279)	(0.0279)	(0.0278)
Constant	-3.298***	-2.245***	-2.658***	-2.931***
	(0.135)	(0.0821)	(0.0857)	(0.131)
Pseudo-R^2	0.111	0.111	0.111	0.111

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Variables	*Group 3* Delinquent Loan	*ModelA13*	*ModelA14*	*ModelA15*
ODTI		0.010***		
		(0.0030)		
ODSTI	0.050***		0.050***	
	(0.0009)		(0.0007)	
OLTV	0.006***			0.004***
	(0.0004)			(0.0004)
Education Level	-0.739***	-0.772***	-0.733***	-0.746***
	(0.0097)	(0.0097)	(0.0097)	(0.0096)
Current Interest Rate	0.523***	0.546***	0.521***	0.543***
	(0.0068)	(0.0067)	(0.0069)	(0.0067)
Number of Debtors	-0.542***	-0.355***	-0.538***	-0.386***
	(0.0166)	(0.0167)	(0.0166)	(0.0168)
Current Age of Main Borrower	-0.015***	-0.005***	-0.016***	-0.011***
	(0.0009)	(0.0010)	(0.0009)	(0.0009)
Loan Age	0.018***	0.018***	0.018***	0.018***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.007***	-0.006***	-0.007***	-0.006***
	(0.0019)	(0.0019)	(0.0019)	(0.0019)
P.Blue Collar Worker	reference			
P.Director/Business Owner	0.703***	0.572***	0.702***	0.697***
	(0.0222)	(0.0225)	(0.0223)	(0.0221)
P.Employee	0.167***	-0.004	0.181***	0.066***
	(0.0251)	(0.0252)	(0.0251)	(0.0250)
P.Employee in Private Sector	0.433***	0.436***	0.429***	0.453***
	(0.0203)	(0.0202)	(0.0203)	(0.0202)
P.Other	0.038	0.009	0.032	0.074*
	(0.0447)	(0.0446)	(0.0447)	(0.0446)
P.Public Servant	-0.093***	-0.076***	-0.097***	-0.078***
	(0.0292)	(0.0292)	(0.0293)	(0.0292)
P.Senior and Medium Manager	0.492***	0.464***	0.503***	0.505***
	(0.0300)	(0.0300)	(0.0300)	(0.0300)
MR.Married	reference			
MR.Other	-0.083***	-0.182***	-0.074***	-0.146***
	(0.0226)	(0.0231)	(0.0226)	(0.0231)
MR.Single	0.426***	0.343***	0.436***	0.400***
	(0.0200)	(0.0203)	(0.0200)	(0.0202)
Constant	-6.779***	-5.224***	-6.415***	-4.796***
	(0.0733)	(0.0661)	(0.0680)	(0.0680)
Pseudo-R^2	0.110	0.0958	0.109	0.0926

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Variables	*Group 4* Delinquent Loan	*ModelA16*	*ModelA17*	*ModelA18*	*ModelA19* Income =< Median	*ModelA20* Income > Median
ODTI		0.029***				
		(0.0028)				
ODSTI	0.073***		0.073***		0.077***	0.066***
	(0.0006)		(0.0006)		(0.0009)	(0.0008)
OLTV	0.002**			0.002***	0.014***	-0.010***
	(0.0008)			(0.0008)	(0.0011)	(0.0012)
Education Level	-0.610***	-0.640***	-0.610***	-0.631***	-0.599***	-0.666***
	(0.0083)	(0.0082)	(0.0083)	(0.0082)	(0.0123)	(0.0117)
Current Interest Rate	0.605***	0.628***	0.607***	0.619***	0.531***	0.659***
	(0.0062)	(0.0061)	(0.0062)	(0.0062)	(0.0099)	(0.0080)
Number of Debtors	-1.164***	-0.912***	-1.165***	-0.922***	-0.723***	-1.483***
	(0.0150)	(0.0155)	(0.0150)	(0.0154)	(0.0238)	(0.0194)
Current Age of Main Borrower	0.008***	0.013***	0.008***	0.012***	-0.002**	0.017***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0013)	(0.0013)
Loan Age	0.013***	0.013***	0.013***	0.013***	0.016***	0.010***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
GDP_{t-3}	0.009***	0.011***	0.009***	0.011***	0.020***	0.001
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0023)	(0.0023)
P.Blue Collar Worker	reference					
P.Director/Business Owner	0.652***	0.630***	0.646***	0.677***	0.634***	0.448***
	(0.0206)	(0.0207)	(0.0204)	(0.0206)	(0.0272)	(0.0337)
P.Employee	0.269***	0.088***	0.260***	0.125***	0.549***	-0.112***
	(0.0201)	(0.0197)	(0.0197)	(0.0200)	(0.0261)	(0.0338)
P.Employee in Private Sector	-0.162***	-0.150***	-0.163***	-0.142***	-0.101***	-0.257***
	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0233)	(0.0299)
P.Other	-1.106***	-1.117***	-1.110***	-1.099***	-1.652***	-1.019***
	(0.0716)	(0.0714)	(0.0715)	(0.0714)	(0.147)	(0.0839)
P.Public Servant	-0.584***	-0.560***	-0.585***	-0.560***	-0.667***	-0.476***
	(0.0270)	(0.0269)	(0.0270)	(0.0269)	(0.0382)	(0.0399)
P.Senior and Medium Manager	0.748***	0.712***	0.747***	0.723***	0.165***	0.744***
	(0.0212)	(0.0209)	(0.0211)	(0.0209)	(0.0445)	(0.0306)
MR.Married	reference					
MR.Other	-0.369***	-0.368***	-0.368***	-0.367***	-0.114***	-0.461***
	(0.0189)	(0.0194)	(0.0189)	(0.0194)	(0.0288)	(0.0258)
MR.Single	-0.379***	-0.409***	-0.377***	-0.403***	-0.328***	-0.250***
	(0.0172)	(0.0178)	(0.0172)	(0.0178)	(0.0267)	(0.0227)
Constant	-7.242***	-4.418***	-7.087***	-4.372***	-8.810***	-5.441***
	(0.0927)	(0.0608)	(0.0611)	(0.0906)	(0.133)	(0.136)
Pseudo-R^2	0.117	0.0842	0.117	0.0840	0.0991	0.147

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 20: Estimation of Other Model from Section 5.3 Initial Income

Variables	*Low Income* Delinquent Loan	*ModelA21*	*ModelA22*	*ModelA23*
ODTI		0.0008		
		(0.003)		
ODSTI	0.023***		0.024***	
	(0.0006)		(0.0006)	
OLTV	0.009***			0.009***
	(0.0003)			(0.0003)
Education Level	-0.825***	-0.814***	-0.818***	-0.821***
	(0.0108)	(0.0108)	(0.0108)	(0.0108)
Current Interest Rate	0.482***	0.515***	0.500***	0.497***
	(0.0084)	(0.0084)	(0.0084)	(0.0084)
Number of Debtors	-0.449***	-0.392***	-0.468***	-0.373***
	(0.0207)	(0.0206)	(0.0207)	(0.0205)
Current Age of Main Borrower	-0.025***	-0.028***	-0.023***	-0.022***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Loan Age	0.013***	0.013***	0.013***	0.013***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.004*	-0.003*	-0.004**	-0.003
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
P.Blue Collar Worker	reference			
P.Director/Business Owner	-0.405***	-0.396***	-0.461***	-0.336***
	(0.0338)	(0.0338)	(0.0337)	(0.0337)
P.Employee	-0.104***	-0.165***	-0.156***	-0.107***
	(0.0223)	(0.0223)	(0.0222)	(0.0223)
P.Employee in Private Sector	-0.014	-0.006	-0.026	0.009
	(0.0179)	(0.0179)	(0.0179)	(0.0179)
P.Other	-0.185***	-0.226***	-0.240***	-0.172***
	(0.0586)	(0.0585)	(0.0586)	(0.0586)
P.Public Servant	-0.098***	-0.096***	-0.121***	-0.069***
	(0.0254)	(0.0254)	(0.0254)	(0.0254)
P.Senior and Medium Manager	-1.541***	-1.488***	-1.542***	-1.484***
	(0.103)	(0.103)	(0.103)	(0.103)
MR.Married	reference			
MR.Other	0.334***	0.360***	0.357***	0.336***
	(0.0275)	(0.0276)	(0.0276)	(0.0276)
MR.Single	0.0467*	0.115***	0.0867***	0.0726***
	(0.0266)	(0.0267)	(0.0267)	(0.0267)
Constant	-4.966***	-3.665***	-4.255***	-4.432***
	(0.0789)	(0.0739)	(0.0735)	(0.0773)
Pseudo-R^2	0.0673	0.0596	0.0648	0.0624

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Variables	*Middle Income* Delinquent Loan	*ModelA24*	*ModelA25*	*ModelA26*
ODTI		0.090***		
		(0.0029)		
ODSTI	0.047***		0.047***	
	(0.0005)		(0.0005)	
OLTV	0.010***			0.011***
	(0.0003)			(0.0003)
Education Level	-0.841***	-0.898***	-0.830***	-0.880***
	(0.0094)	(0.0094)	(0.0094)	(0.0094)
Current Interest Rate	0.492***	0.525***	0.490***	0.516***
	(0.0071)	(0.0070)	(0.0071)	(0.0071)
Number of Debtors	-0.703***	-0.585***	-0.707***	-0.590***
	(0.0157)	(0.0159)	(0.0157)	(0.0158)
Current Age of Main Borrower	-0.027***	-0.026***	-0.034***	-0.024***
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Loan Age	0.018***	0.019***	0.018***	0.019***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_{t-3}	-0.001	-0.0003	-0.002	-0.0002
	(0.0018)	(0.0018)	(0.0018)	(0.0018)
P.Blue Collar Worker	reference			
P.Director/Business Owner	0.747***	0.536***	0.660***	0.745***
	(0.0245)	(0.0245)	(0.0243)	(0.0243)
P.Employee	0.361***	0.109***	0.302***	0.238***
	(0.0226)	(0.0225)	(0.0225)	(0.0225)
P.Employee in Private Sector	0.200***	0.145***	0.170***	0.207***
	(0.0208)	(0.0208)	(0.0208)	(0.0208)
P.Other	0.495***	0.300***	0.450***	0.401***
	(0.0391)	(0.0390)	(0.0390)	(0.0389)
P.Public Servant	-0.345***	-0.387***	-0.378***	-0.337***
	(0.0328)	(0.0328)	(0.0328)	(0.0328)
P.Senior and Medium Manager	0.381***	0.285***	0.367***	0.367***
	(0.0288)	(0.0288)	(0.0288)	(0.0287)
MR.Married	reference			
MR.Other	0.124***	0.150***	0.165***	0.115***
	(0.0206)	(0.0209)	(0.0206)	(0.0209)
MR.Single	0.044**	0.060***	0.087***	0.023
	(0.0183)	(0.0185)	(0.0183)	(0.0185)
Constant	-5.945***	-3.936***	-4.992***	-4.376***
	(0.0748)	(0.0675)	(0.0671)	(0.0723)
Pseudo-R^2	0.104	0.0782	0.102	0.0785

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Variables	*High Income* Delinquent Loan	*ModelA27*	*ModelA28*	*ModelA29*
ODTI		0.258***		
		(0.0036)		
ODSTI	0.061***		0.060***	
	(0.0008)		(0.0007)	
OLTV	0.008***			0.007***
	(0.0005)			(0.0005)
Education Level	-0.592***	-0.708***	-0.571***	-0.615***
	(0.0149)	(0.0149)	(0.0149)	(0.0146)
Current Interest Rate	0.535***	0.542***	0.519***	0.550***
	(0.0089)	(0.0088)	(0.0089)	(0.0088)
Number of Debtors	-1.567***	-1.514***	-1.565***	-1.570***
	(0.0242)	(0.0245)	(0.0242)	(0.0248)
Current Age of Main Borrower	0.030***	0.049***	0.025***	0.034***
	(0.0015)	(0.0015)	(0.0014)	(0.0015)
Loan Age	0.019***	0.019***	0.019***	0.020***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
GDP_{t-3}	-0.001	-0.0001	-0.002	0.0009
	(0.0028)	(0.0028)	(0.0028)	(0.0028)
P.Blue Collar Worker	reference			
P.Director/Business Owner	1.342***	1.040***	1.287***	1.480***
	(0.133)	(0.133)	(0.133)	(0.133)
P.Employee	0.703***	0.412***	0.678***	0.612***
	(0.134)	(0.135)	(0.134)	(0.134)
P.Employee in Private Sector	0.556***	0.550***	0.559***	0.563***
	(0.135)	(0.135)	(0.135)	(0.135)
P.Other	0.621***	0.466***	0.587***	0.704***
	(0.141)	(0.141)	(0.141)	(0.141)
P.Public Servant	0.506***	0.762***	0.524***	0.628***
	(0.138)	(0.139)	(0.138)	(0.138)
P.Senior and Medium Manager	1.099***	0.999***	1.109***	1.045***
	(0.133)	(0.134)	(0.133)	(0.133)
MR.Married	reference			
MR.Other	-0.573***	-0.667***	-0.564***	-0.674***
	(0.0333)	(0.0334)	(0.0333)	(0.0336)
MR.Single	0.214***	0.017	0.231***	0.085***
	(0.0276)	(0.0281)	(0.0276)	(0.0280)
Constant	-8.363***	-7.064***	-7.620***	-6.113***
	(0.171)	(0.164)	(0.163)	(0.167)
Pseudo-R^2	0.194	0.179	0.192	0.147

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Estimation of Other Model from Section 5.4 Economic Sensitivity

Variables	*ModelA30* Delinquent Loan	*ModelA31*	*ModelA32*	*ModelA33*	*ModelA34*	*ModelA35*
ODTI	0.0985***			0.0985***		
	(0.00215)			(0.00215)		
ODSTI		0.0453***			0.0453***	
		(0.000398)			(0.000398)	
OLTV			0.0110***			0.0110***
			(0.000249)			(0.000249)
Education Level	-0.635***	-0.599***	-0.629***	-0.635***	-0.600***	-0.629***
	(0.00712)	(0.00712)	(0.00708)	(0.00712)	(0.00713)	(0.00708)
Current Interest Rate	0.501***	0.465***	0.486***	0.497***	0.466***	0.481***
	(0.00535)	(0.00541)	(0.00539)	(0.00551)	(0.00555)	(0.00557)
Number of Debtors	-0.674***	-0.798***	-0.678***	-0.675***	-0.798***	-0.678***
	(0.0131)	(0.0129)	(0.0131)	(0.0131)	(0.0129)	(0.0131)
Current Age of Main Borrower	-0.00369***	-0.0138***	-0.00345***	-0.00354***	-0.0138***	-0.00325***
	(0.000742)	(0.000720)	(0.000739)	(0.000744)	(0.000722)	(0.000740)
Loan Age	0.0149***	0.0151***	0.0152***	0.0154***	0.0152***	0.0157***
	(0.000166)	(0.000165)	(0.000166)	(0.000189)	(0.000188)	(0.000189)
c GDP_{t-6}	-0.0221***	-0.0231***	-0.0218***			
	(0.00128)	(0.00129)	(0.00128)			
$Unemployment_{t-3}$				0.00405	-0.0171***	0.00731
				(0.00519)	(0.00518)	(0.00521)
P.Blue Collar Worker	reference					
P.Director/Business Owner	0.416***	0.445***	0.602***	0.425***	0.447***	0.612***
	(0.0175)	(0.0173)	(0.0174)	(0.0176)	(0.0174)	(0.0175)
P.Employee	-0.0446***	0.0575***	0.0619***	-0.0145	0.0631***	0.0958***
	(0.0163)	(0.0163)	(0.0163)	(0.0174)	(0.0174)	(0.0174)
P.Employee in Private Sector	-0.0952***	-0.110***	-0.0553***	-0.0972***	-0.109***	-0.0575***
	(0.0152)	(0.0152)	(0.0152)	(0.0152)	(0.0152)	(0.0152)
P.Other	0.195***	0.217***	0.288***	0.202***	0.219***	0.295***
	(0.0305)	(0.0305)	(0.0305)	(0.0305)	(0.0305)	(0.0305)
P.Public Servant	-0.298***	-0.336***	-0.262***	-0.300***	-0.335***	-0.265***
	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)
P.Senior and Medium Manager	0.277***	0.277***	0.309***	0.275***	0.278***	0.307***
	(0.0212)	(0.0212)	(0.0212)	(0.0212)	(0.0212)	(0.0212)
MR.Married	reference					
MR.Other	-0.0689***	-0.0128	-0.0740***	-0.0690***	-0.0123	-0.0742***
	(0.0168)	(0.0165)	(0.0167)	(0.0168)	(0.0165)	(0.0167)
MR.Single	-0.0742***	-0.0273*	-0.0797***	-0.0720***	-0.0280*	-0.0773***
	(0.0155)	(0.0152)	(0.0155)	(0.0155)	(0.0153)	(0.0155)
Constant	-5.167***	-5.864***	-5.484***	-5.247***	-5.859***	-5.577***
	(0.0496)	(0.0489)	(0.0525)	(0.0539)	(0.0530)	(0.0567)
Pseudo-R^2	0.0624	0.0803	0.0624	0.0619	0.0798	0.0620

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 22: Estimation of Random Effects Panel Logit Model (Section 5.6)

Variables	*RE panel logit* Delinquent Loan
ODSTI	0.042***
	(0.0020)
OLTV	0.014***
	(0.0012)
Education Level	-0.189***
	(0.0317)
Current Interest Rate	0.034**
	(0.0165)
Number of Debtors	-0.710***
	(0.0534)
Current Age of Main Borrower	0.020***
	(0.0037)
GDP_{t-3}	-0.0144***
	(0.0020)
P.Blue Collar Worker	reference
P.Director/Business Owner	0.292***
	(0.100)
P.Employee	0.961***
	(0.0847)
P.Employee in Private Sector	-0.951***
	(0.0817)
P.Other	0.356**
	(0.171)
P.Public Servant	-0.890***
	(0.122)
P.Senior and Medium Manager	-0.490***
	(0.110)
MR.Married	reference
MR.Other	-0.025
	(0.0849)
MR.Single	0.009
	(0.0690)
Loan Age	0.041***
	(0.0005)
$/\ln(\sigma_u^2)$	3.809***
	(0.0056)
Constant	-23.63***
	(0.239)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 23: Estimation of Cluster-Based Logit Model (Section 5.6)

Variables	*Cluster-based logit* Delinquent Loan
ODSTI	0.047***
	(0.0025)
OLTV	0.012***
	(0.0014)
Education Level	-0.471***
	(0.0424)
Current Interest Rate	0.394***
	(0.0263)
Number of Debtors	-0.785***
	(0.0946)
Current Age of Main Borrower	-0.004
	(0.0045)
GDP_{t-3}	-0.013***
	(0.0033)
P.Blue Collar Worker	reference
P.Director/Business Owner	0.456***
	(0.105)
P.Employee	0.126
	(0.0901)
P.Employee in Private Sector	-0.262***
	(0.0943)
P.Other	0.264
	(0.161)
P.Public Servant	-0.405***
	(0.132)
P.Senior and Medium Manager	0.0195
	(0.131)
MR.Married	reference
MR.Other	-0.078
	(0.104)
MR.Single	-0.099
	(0.102)
Loan Age	0.015***
	(0.0007)
Constant	-7.462***
	(0.331)
Pseudo-R^2	0.0709

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24: Estimation of Cross-Section Logit Model (Section 5.6)

Variables	*Cross-section analysis* Delinquent Loan
ODSTI	0.044***
	(0.0015)
OLTV	0.007***
	(0.0009)
Education Level	-0.592***
	(0.0253)
Current Interest Rate	0.403***
	(0.0225)
Number of Debtors	-0.679***
	(0.0456)
Current Age of Main Borrower	-0.001
	(0.0026)
Loan Age	0.052***
	(0.0016)
GDP_{t-3}	-0.092***
	(0.0048)
P.Blue Collar Worker	reference
P.Director/Business Owner	0.127*
	(0.0665)
P.Employee	-0.484***
	(0.0592)
P.Employee in Private Sector	-0.151***
	(0.0577)
P.Other	0.052
	(0.107)
P.Public Servant	-0.220***
	(0.0775)
P.Senior and Medium Manager	0.319***
	(0.0766)
MR.Married	reference
MR.Other	0.064
	(0.0588)
MR.Single	0.006
	(0.0541)
Constant	-6.104***
	(0.207)
Pseudo-R^2	0.136

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25: Estimation of BMA-Selected Logit Model (Section 5.6)

Variables	*BMA selection* Delinquent Loan
ODSTI	0.045***
	(0.0004)
OLTV	0.010***
	(0.0002)
Education Level	-0.616***
	(0.0064)
Current Interest Rate	0.468***
	(0.0055)
Number of Debtors	-0.753***
	(0.0098)
Loan Age	0.014***
	(0.0001)
GDP_{t-3}	-0.010***
	(0.0013)
Constant	-7.034***
	(0.0370)
Pseudo-R^2	0.0793

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1